DOCUMENT RESUME

ED 350 340 TM 019 101

TITLE	Potential Strategies for Long-Term Indicator Development. Reports of the Technical Planning Subgroups.
INSTITUTION	National Education Goals Panel, Washington, DC.
REPORT NO	NEGP-91-08
PUB DATE	4 Sep 91
NOTE	127p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Academic Achievement; *Academic Standards; Adult Literacy; Early Childhood Education; Educational Assessment; Educational Change; Educational Improvement; *Educational Objectives; Elementary Secondary Education; *Guidelines; High School Graduates; *Long Range Planning; Mathematics Achievement; National Competency Tests; School Restructuring; Science Education
IDENTIFIERS	America 2000; *Educational Indicators; *National Education Goals 1990; Progress Reporting; Reform Efforts

ABSTRACT

The National Education Goals Panel was created by the President and Governors in July 1990 to assess and report on the progress of the nation and states in achieving six National Education goals by the year 2000. The Panel created six resource groups, one for each goal, to develop short- and long-term recommendations for methods to ensure progress. The recommendations of the resource groups were forwarded to Technical Planning Subgroups that were created to address the following topics: (1) Goal 1 (all children will start school ready to learn), developing an early childhood assessment system; (2) Goal 2 (increasing the high school graduation rate to at least 90 percent), developing a national student records system; (3) Goal 3 (competence in challenging subject matter), developing a national system of examinations linked to common standards; (4) Goal 4 (U.S. students will be first in the world in science and mathematics), creating improved indicators of science and mathematics instruction; (5) Goal 5 (every adult will be literate), obtaining international workforce comparisons and developing examination systems; and (6) Goal 6 (schools will be safe, disciplined, and drug-free), creating improved state and national indicators of such schools. This volume contains the recommendations of the resource groups and the technical planning subgroups, goal by goal. An appendix contains a summary of preliminary work of the subgroups. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED350340

NATIONAL EDUCATION GOALS PANEL

POTENTIAL STRATEGIES FOR LONG-TERM INDICATOR DEVELOPMENT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

Reports of the Technical Planning Subgroups

September 4, 1991

91-08

BEST COPY AVAILABLE

TM019101

National Education Goals Panel

MEMBERS

GOVERNORS

Carroll A. Campbell, Jr., South Carolina, Chairman
Roy Romer, Colorado, Past Chairman
John Ashcroft, Missouri
Evan Bayh, Indiana
Terry E. Branstad, Iowa
Booth Gardner, Washington

MEMBERS OF THE ADMINISTRATION

Secretary of Education Lamar Alexander
The Director of Office Management and Budget Richard Darman
Assistant to the President for Economic
and Domestic Policy Roger B. Porter
White House Chief of Staff Governor John Sununu

MEMBERS OF CONGRESS (ex officio)

U.S. House Majority Leader, Richard Gephardt
U.S. House Minority Leader, Robert Michel
U.S. Senate Majority Leader, George J. Mitchell
U.S. Senate Minority Leader, Bob Dole

Dr. Pascal D. Forgione, Jr.
Executive Director

TABLE OF CONTENTS

EXPERT RECOMMENDATIONS FROM THE FIELD

Introduction ii

Chapter 1: Readiness for School 1
Recommendations from the Goal 1 Resource Group and Technical Planning Subgroup

Chapter 2: High School Completion 19
Recommendations from the Goal 2 Resource Group and Technical Planning Subgroup

Chapter 3: Student Achievement and Citizenship 41
Recommendations from the Goal 3 Resource Group and Technical Planning Subgroup

Chapter 4: Science and Mathematics 65
Recommendations from the Goal 4 Resource Group and Technical Planning Subgroup

Chapter 5: Adult Literacy and Lifelong Learning 79
Recommendations from the Goal 5 Resource Group and Technical Planning Subgroup

Chapter 6: Safe, Disciplined, and Drug-free Schools 101
Recommendations from the Goal 6 Resource Group and Technical Planning Subgroup

Appendix A: 115
Preliminary work of the Technical Planning Subgroups on Citizenship and The National Assessment of Educational Progress (NAEP)

PUBLIC COMMENT

Information Request Form
Public Comment Form

EXPERT RECOMMENDATIONS FROM THE FIELD

INTRODUCTION

Background

The National Education Goals Panel was created by the President and Governors in July 1990 to assess and report on the progress of the nation and states toward achieving six National Education Goals by the year 2000:

Goal 1 – all children will start school ready to learn;

Goal 2 – the high school graduation rate will increase to at least 90 percent;

Goal 3 – students will demonstrate competence in challenging subject matter;

Goal 4 – U.S. students will be first in the world in science and mathematics;

Goal 5 – every adult will be literate and have the skills and knowledge to compete in a global economy and function as citizens; and

Goal 6 – schools will be safe, disciplined and drug-free.

The Panel is an independent, bipartisan fourteen member group charged with providing parents, educators, and policymakers with information on national and state performance in each goal area. It officially began its work in August 1990 under the chairmanship of Colorado Governor Roy Romer. As part of its operating principles, the Panel agreed at its first meeting in October 1990 to seek the advice and counsel of education and assessment experts and to build upon exemplary assessment practices at the national and state level. In February 1991, Pascal D. Forgione, Jr. was appointed Executive Director of the National Education Goals Panel and an independent office was established to assist and to advise the Panel on the selection of goal indicators and assessment strategies for measuring progress toward achieving the National Education Goals.

The Panel created six Resource Groups (one for each Goal) to develop both short and long term recommendations for the Panel's review and possible adoption of methods to measure progress toward the Goals. For the short-term, the Resource Groups were asked to identify and recommend currently available data sources that could be used to report progress in the first annual Goals Report issued in September 1991. For the long-term, the Resource Groups were asked to identify the major gaps in currently available data for determining whether the Goals were being achieved.

The Resource Groups presented interim reports to the Panel in March 1991 in a two-volume publication entitled Measuring Progress Toward the National Education Goals: Potential Indicators and Measurement Strategies. Volume I, Discussion Document, provides an executive summary of each Resource Groups preliminary recommendations; Volume II, Compendium of Resource Group Interim Reports, contains the full text of each Resource Group's report.

In June and July of 1991 the Panel forwarded the long-term recommendations of the six Resource Groups to their counterpart Technical Planning Subgroups. The Technical Planning Subgroups were created to provide additional details on selected Resource Group proposals for developing new indicators and assessment strategies to report on progress in future Goals Reports. The Subgroups were created to address the following topics:

Goal 1 – developing an early childhood assessment system;

Goal 2 – developing a national student records system;

Goal 3 – developing a national system of examinations linked to common standards;

Goal 4 – creating improved indicators of science and mathematics instructional practices;

Goal 5 – obtaining international workforce comparisons and developing national and collegiate examination systems; and

Goal 6 – creating improved state and national indicators of safe, disciplined and drug-free schools.

In August, each of these Subgroups submitted a report to their respective Resource Group. The Resource Groups then transmitted the reports and their comments to the Goals Panel.

The Technical Planning Subgroup reports were presented to the Panel at a September 4th meeting chaired by South Carolina Governor Carroll Campbell who assumed the position of Panel Chairman in August 1991. The convener of each Resource Group introduced each report and the leader of each Technical Planning Subgroup discussed the recommendations for future indicators and measurement strategies.

This volume contains the recommendations of the six Resource Groups and their respective Technical Planning Subgroups. It first presents, within each Goal area, a memorandum from the Resource Group summarizing and commenting on the Subgroup's work. This is followed by the full text of the Technical Planning Subgroup Report.

The Goals Panel is seeking broad reaction to the Technical Planning documents from interested parties. A comment form is included at the end of this volume to facilitate feedback.

CHAPTER 1

GOAL 1: READINESS FOR SCHOOL

GOAL 1: By the year 2000, all children in America will start school ready to learn.

Objectives:

- All disadvantaged and disabled children will have access to high quality and developmentally appropriate preschool programs that help prepare children for school.
- Every parent in America will be a child's first teacher and devote time each day helping his or her preschool child learn; parents will have access to the training support they need.
- Children will receive the nutrition and health care needed to arrive at school with healthy minds and bodies, and the number of low birth weight babies will be significantly reduced through enhanced prenatal health systems.

GOAL 1 RESOURCE GROUP ON READINESS FOR SCHOOL

RESOURCE GROUP MEMBERSHIP

Convener	
Ernest L. Boyer	The Carnegie Foundation for the Advancement of Teaching, New Jersey
Members	
James P. Comer	Yale University, Child Studies Center, Connecticut
Donna Foglia	Evergreen School District, California
Sharon Lynn Kagan	Yale University, The Bush Center in Child Development and Social Policy, Connecticut
Samuel Meisels	University of Michigan, Center for Human Growth and Development
Lucile F. Newman	Brown University, School of Medicine, Rhode Island
Douglas Powell	Purdue University, Child Development and Family Studies, Indiana
James Wilsford	Orangeburg School District, South Carolina
Nicholas Zill	Child Trends, Inc., District of Columbia

GOAL 1 TECHNICAL PLANNING SUPBGROUP ON READINESS FOR SCHOOL

TECHNICAL PLANNING SUBGROUP MEMBERSHIP

Leader	
Sharon Lynn Kagan	Yale University, The Bush Center in Child Development and Social Policy, Connecticut
Members	
Ernest L. Boyer	The Carnegie Foundation for the Advancement of Teaching, New Jersey
Sue Bradekamp	National Association for the Education of Young Children, District of Columbia
M. Elizabeth Graue	University of Wisconsin – Madison
Luis Laosa	Educational Testing Service, New Jersey
Samuel Meisels	University of Michigan, Center for Human Growth and Development
Lucile F. Newman	Brown University, School of Medicine, Rhode Island
Lorrie Shepard	University of Colorado, School of Education
Joseph Stevens	Georgia State University, College of Education
Sheldon White	Harvard University, Department of Psychology, Massachusetts
Nicholas Zill	Child Trends, Inc., District of Columbia

September 4, 1991

TO: THE NATIONAL EDUCATION GOALS PANEL

FROM: ERNEST BOYER, CONVENER
THE RESOURCE GROUP ON SCHOOL READINESS

GOAL 1 RESOURCE GROUP

STATEMENT ON THE TECHNICAL PLANNING SUBGROUP'S REPORT

The Goal 1 Resource Group is pleased to submit to the National Education Goals Panel the final report of its Technical Planning Subgroup on the feasibility and viability of a national assessment of kindergarten students. We urge all who are concerned to read the full report for its careful consideration of the many difficult technical, educational, and ethical issues involved.

The Resource Group strongly endorses the Technical Planning Subgroup's report on the development of a system of assessment, which would provide comprehensive information about the status of the nation's children as they enter school. Rather than a single measure or index of "readiness," the technical report recommends an early childhood assessment designed to draw a profile of kindergarten children along several dimensions of early learning and development, from as many perspectives as possible. Through parent reports, teacher reports, performance portfolios, and a profile of children's skills, knowledge, and development, this assessment would describe five characteristics that enable children to take advantage of the opportunities and demands of formal schooling. As defined in the technical report, these characteristics include physical well–being and motor development; social and emotional development; approaches toward learning; language usage; and cognition and general knowledge.

The Technical Subgroup sees the development of this system of assessment as an opportunity not only to advance a truly holistic definition of school readiness, but also to expand knowledge about how to assess young children equitably, in ways that do not label, stigmatize, or classify them. Because the purpose of the early childhood assessment is to provide a national overview of young children's early learning and development, rather than an assessment of individual children or groups of children, the subgroup recommends that both children and assessment items be sampled. Further, given the complexity of the assessment tasks, the subgroup also suggests that data collection occur not annually, but every three years.

The Resource Group wishes, especially, to support the Subgroup's recommendation for the formation of a National Commission on Early Childhood Assessment to supervise the development and implementation of such an assessment, and later, to evaluate the assessment itself. Difficult technical issues will have to be addressed in order to ensure valid and reliable means of assessing the various dimensions of early learning and development of the nation's children, and it is of critical importance that the multidimensional approach to early childhood assessment be preserved. For this reason, the National Commission on Early Childhood Assessment should receive a long-term commitment, to ensure that the system of assessment, once in place continues to meet its objectives by providing high quality data to assist the nation as it strives to improve services and outcomes for young children. If the Panel decides to endorse an in-school assessment of young children, we urge that efforts to fund and staff the commission begin as soon as possible.

Finally, in regard to the larger task of reporting to the nation on Goal 1, there are other steps to be taken, too. It is important to recall that the Goal 1 Resource Group recommended monitoring children' early progress at three points in time. In addition to in-school assessment, we recommended the collection of information about children's health, home life, and preschool experience at the time of school entry, and also before they enter school. We wish to restate here our conviction that all three are critically important, and we urge the National Education Goals Panel to request further examination of school-entry and before-school data, possibly as a task for the Resource Group itself.

THE GOAL 1 TECHNICAL PLANNING SUBGROUP REPORT ON SCHOOL READINESS

Submitted to

The National Education Goals Panel

through

THE GOAL 1 RESOURCE GROUP ON SCHOOL READINESS

for presentation at

The National Education Goals Panel Meeting

September 4, 1991

THE GOAL 1 TECHNICAL PLANNING SUBGROUP REPORT ON SCHOOL READINESS

INTRODUCTION

In its report to the National Education Goals Panel, the Goal 1 Resource Group recommended monitoring children's early progress at three points: 1) before school; 2) upon entry to school; and 3) in school. The Resource Group formed a Technical Subgroup to examine the issues associated with the development and implementation of the in-school assessment. Specifically, the Technical Subgroup was asked to assess the feasibility and viability of developing a direct measure of children's readiness.

In chronicling the Technical Subgroup's recommendations, this report is premised on the strong belief that a comprehensive, well-conceived, and well-implemented system of assessment, in direct contrast to a single measure or index, could be helpful in providing information regarding children's potential for succeeding in school and in fostering the improvement of children's school experiences. Indeed, the Technical Subgroup regards the development of the entire assessment system discussed herein as a rich opportunity to advance a holistic definition of readiness, to further the nation's equitable assessment of all children, and to develop greater understanding of the condition of the nation's young children as a prelude to improving the services that foster children's development and learning.

In offering recommendations, we have sought to pay close attention to nomenclature, knowing full well that the name given to an assessment program and the dimensions selected for measurement are of profoundly symbolic importance and might affect local and state assessment practices and policies. Because current practice contains so much evidence of the potential misuse of readiness tests and of the misapplication of developmental screening tests, we have assiduously avoided the use of the term "readiness" in this document, and have attempted to devise an assessment process that will have as few negative consequences for individual children as possible. Thus, we suggest that the National Education Goals Panel endorse the development of an early childhood assessment system to monitor the nation's progress on Goal 1. We further urge the Panel to ensure that no single index or instrument be used, and that the assessment system to be created not label, stigmatize, or classify any individual child or group of children.

To that end, the **Early Childhood Assessment** that we propose is designed to provide comprehensive information about the status of the nation's children during their kindergarten year. It is based on four fundamental assumptions:

1. Being prepared to participate successfully in formal schooling is a lengthy, complex, and multidetermined process, reflecting the complex nature of child growth and development and the multiple forces that shape young children's lives. Thus, an adequate assessment of young children's preparedness for formal schooling must incorporate all facets of the growing child, and must utilize information from a variety of sources.

2. Given the multidimensional conception of early school experience that we recommend, an adequate assessment program must include multiple assessment perspectives. No single instrument or questionnaire, nor any other single source of data, can capture this complex view.

3. The purpose of the Early Childhood Assessment is to provide a national overview of young children's early learning and development, rather than an assessment of individual children or groups of children. Thus, for practical reasons it is essential that children be sampled. We do not wish to propose an assessment mechanism that can potentially disadvantage or stigmatize any individual child or group of children.

4. Although the Early Childhood Assessment is focused on children during their kindergarten year, in accord with the Resource Group's recommendation, data should also be obtained about children's experiences and status prior to formal school entry. Indicators of prenatal health, parental age and education, preschool program participation, preschoolers' activities in the home, children's access to health care, quality of preschool programs, and nutritional status are among the types of indicators that will enrich our understanding of the resources and opportunities that children bring to school.

In short, development of this kind of assessment represents a new direction for the field of early childhood. Previous assessment work has generally addressed the need to compare individual children in relationship to their peers; to determine if a child was delayed developmentally; to make instructional decisions; to assess achievement; or to determine special educational placements. The Early Childhood Assessment is not a "test" in the sense implied by the above examples. Rather, it is a means of obtaining data on the status of the nation's children in order to inform public policy. Such information can be useful to policy makers in charting progress toward achievement of the National Educational Goals, and for informing the development, expansion, and/or modification of policies and programs that affect young children and their families.

The shift to a large-scale assessment model such as the Early Childhood Assessment, with its comprehensive focus and its ambition to provide a national early childhood status report, will require the development of new assessment strategies and instruments for use with young children.

DIMENSIONS OF LEARNING AND DEVELOPMENT

Children's preparedness for school is typically assessed in a one-dimensional manner by the administration of a test that narrowly evaluates a child's general knowledge and abilities at the outset of school. We reject this reductionistic approach as one that cannot possibly encompass the range of abilities and experiences brought by children when they enter school, nor the range of knowledge and skills that they will need to be successful once they arrive there. In contrast, we propose a conception of early learning and development that is multifaceted, reflecting a variety of elements relating to children's status at the outset of school. Such elements or dimensions concern a child's physical well-being, social and emotional development, approaches to learning, use of language, and cognitive skills and general knowledge. These dimensions of early learning and development describe the whole child, rather than a single element of the child that may be related to a specific outcome. The dimensions can be further elaborated as follows:

1. **Physical Well-Being and Motor Development.** A strong body of research evidence links maternal and child health to performance in school. We know that conditions such as very low birthweight and poor nutrition may have long-term effects on a child's preparedness for school. Basic information about the child's health history is vital for understanding how children come to school. In addition, early childhood educators emphasize the importance of optimal motor development in children, ranging from the large motor movements that occur on the playground to the small motor work required for holding a crayon or putting together puzzles.

2. **Social and Emotional Development.** This dimension serves as the foundation for relationships that give meaning to school experience. It involves a sense of personal well-being that comes from stable interactions in the child's early life, and interactions that enable children to participate in classroom activities that are positive for themselves, their classmates, and their teachers. Critically important conditions of social and emotional development include emotional support and secure relationships which engender the child's acquisition of such characteristics as self confidence and the ability to function as a member of a group (e.g., taking turns, following directions, and performing tasks which contribute to a creative and productive classroom climate).

3. **Approaches toward Learning.** By "approaches to learning" we are referring to the inclinations, dispositions, or styles -- rather than skills -- that reflect the myriad ways that children become involved in learning, and develop their inclinations to pursue it. Approaches to learning that vary within and between cultures must be respected so we do not advocate a uniform or "cookie cutter" approach to early childhood education, with the goal of all children coming out the same. A child can be successful in school in many ways, and families and teachers need to understand the various ways that children become engaged in learning in order to know how to

enhance and not discourage their engagement. Curiosity, creativity, independence, cooperativeness, and persistence are some of the kinds of approaches that enhance early learning and development.

4. **Language Usage.** Language empowers children to participate in both the cognitive and affective components of the educational program. Experience with language, in both written and oral form, provides children with the tools to interact with others and to represent their thoughts, feelings, and experiences. Communicating effectively with other children and adults and having emergent literacy experiences with the many forms of language are fundamental aspects of this dimension.

5. **Cognition and General Knowledge.** A foundation for later learning is provided when children have opportunities to interact with others and materials and, as a result, are encouraged to learn from their surroundings. Children's transitions to formal schooling are eased when they have been provided with a variety of play–oriented exploratory activities, and when their early school experiences continue these activities. Cognition and general knowledge represent the accumulation and reorganization of experiences that result from participating in a rich learning setting with skilled and appropriate adult intervention. From these experiences children construct knowledge of patterns and relations, cause and effect, and methods of solving problems in everyday life.

Clearly, this definition of early learning and development is neither simple nor narrow. It conveys a pattern of qualities, or a cluster of conditions and characteristics, that allow children to take advantage of the opportunities and demands of formal schooling. Attending to all five of these dimensions is basic to any early childhood program or kindergarten; no one dimension should be stressed more than another.

It should be emphasized, however, that these dimensions cannot be aggregated in any way to form a single quotient, index, or score. Children are not "ready" or "unready," "advanced" or "delayed," "normal" or "abnormal." Rather, they display patterns of skill and experience on continuums of physical well–being, social and emotional development, approaches to learning, language usage, and cognition and general knowledge. Relative strengths and weaknesses on any of these dimensions limit or enhance early school experience. We hope that this comprehensive perspective will enhance the likelihood of finding ways to enrich children's experiences so that all children will experience school success.

THE EARLY CHILDHOOD ASSESSMENT

We propose an Early Childhood Assessment that collects information on the above dimensions from a variety of sources, including parents, other primary caregivers, teachers and individual children. Conceptualized as an interlocking system, the Early Childhood

Assessment will provide rich information about children and their early experiences from multiple perspectives at various points in time. It will allow us to look back on children's

preschool years, helping us to identify circumstances and conditions that have an impact on children's learning and development. It will also allow a look forward to anticipate the skills and resources that children might need to be successful educationally. However, the Early Childhood Assessment is not intended to be predictive in the traditional sense of the word. Rather, it should enable us to assess the kinds of skills and experiences that we expect all children will have when they are of kindergarten age, and it should clarify the variety of challenges that children will present at the outset of formal schooling.

The Early Childhood Assessment should be thought of as a large scale assessment system to provide information for the aggregate population. Within this system we propose a national probability sample of children in both public and private kindergarten settings. We suggest sampling students because we want to assure that the proposed assessment will not become a high-stakes test used for making classification decisions about individual students. In addition, we are concerned that any instrument that samples all of the information we require for this program may be burdensome to an individual child, family, and teacher. Therefore, to avoid using assessment results for classifying children, to guarantee confidentiality, and to minimize burden we strongly recommend sampling of children. We also recommend that consideration be given to sampling items. Spreading meaningful subsets of items across children may provide the broad view that we are seeking while maintaining sufficient validity to support policy generalizations.

In recommending that children be assessed during their kindergarten year, we are aware that there is not a uniform age across the country for kindergarten entry and that not all children attend formal educational programs before first grade. Consequently, devising an inclusive nationally representative sample of children prior to first grade is a complex issue that warrants careful attention during the implementation phase.

SOURCES OF INFORMATION

The proposed assessment has four components or information sources: parent reports, teacher reports, a profile of children's skills, knowledge, and development, and performance portfolios of children's work. By attending to all dimensions of child development from a variety of perspectives, we expect to obtain a richer, more reliable view of the resources available to children as they begin formal schooling.

Parent Reports. The Early Childhood Assessment places a high priority on the value of parents and other primary caregivers as excellent sources of information. They can be a potential source of health information, when there has been continuity of care, concerning conditions of birth, immunizations, disabilities, and other health factors that might affect

school attendance and learning (e.g., severe disease, childhood accidents, or chronic illness). They can also provide insight into child behavior, interaction, and skills as seen from the perspective of the home -- the child's first learning setting. Focus will be on the activities in which children engage at home and the types of skills that they exhibit in those activities. In addition, background questions will provide information on issues of access to such early childhood services as screening, health and nutrition programs, and preschool and child care programs.

Teacher Reports. Teachers collect a wealth of information as they interact with children in the classroom setting. The Teacher Report will provide a systematic opportunity for teachers to provide information gathered from large and small group activity, one-to-one interaction, and informal observation of children in the classroom. The Teacher Report should be completed after the teacher has had time to get to know the child fairly well; a suggested administration point is at the beginning of the second semester of kindergarten. A checklist format is proposed in order to simplify both administration and scoring. In addition to items about individual children's activities in the classroom, and their early school experiences, the teacher will be asked to reply to descriptive questions about the range of children with whom they work, their curriculum goals, and their own educational backgrounds. This information will help us to gain a better understanding of the context of classroom practice.

Profile of Children's Skills, Knowledge, and Development. This source of information will take the form of an individually-administered inventory that samples behavior at a single point in time, providing information for a national profile rather than an evaluation of a single child. It is a snapshot of the skills of children at the midpoint of their kindergarten year. It will focus on such developmental abilities as cognitive processes, reasoning, and adaptive motor functioning as well as skills and general knowledge relevant to the development of literacy, numeracy, and problem-solving. Included will be items intended to assess language skills, listening skills (e.g., retelling a story), familiarity with print (e.g., awareness that books contain stories and that it is possible to learn to read books), counting (e.g., one-to-one correspondence), social skills for interacting with adults and other children, and general knowledge about the environment, the community, and the natural world. In designing this component, it will be essential that the content is broad enough and desirable enough that negative instructional influences would be minimized if the items were misused to shape curriculum in individual classrooms.

Performance Portfolios. The Portfolio attempts to capture the richness of educational products collected over time. A complement to the single-point snapshot provided by the Profile of the Children's Skills, Knowledge, and Development, the Portfolio provides a sample of the products that emerge from educational activities, allowing a qualitative evaluation of children's skills and interests. Portfolios give a highly contextualized view of child activity, more clearly addressing the child's work in a particular educational setting, and providing unique information that cannot be derived from any of the other tools we have suggested. In the Early Childhood Assessment, portfolio collection would occur at several time points in the middle of the kindergarten year.

DEVELOPMENT ISSUES FOR THE MEASUREMENT COMPONENT

General Methodological Concerns

1. ***Although all assessment components are important, their contributions to our knowledge are not the same across all dimensions.***
Working from the assumption that all five dimensions are crucial to our complete understanding of children's early learning and development, nevertheless, some of the tools will be more useful than others in providing information about certain dimensions. For example, the Performance Portfolios will probably contribute more to our understanding of general knowledge than they will to knowledge of the child's physical well-being. This is a question of utility, in that it focuses on which information can best be obtained from the proposed assessment tools.

2. ***Technical expertise varies by dimension and assessment component.***
A second issue concerns the current level of technical measurement expertise. The potential utility of some of the assessment components is not presently within our reach because our technical knowledge base is not sufficiently well-developed. For example, the Early Childhood Assessment must address social and emotional development, but our ability to incorporate this area into the assessment system is limited by our experience and technical knowledge at this point in time. The focus here should be on the development of measurement expertise in particular areas, specifically the assessment of social and emotional development in young children.

3. ***The impact of the assessment on early childhood practices must be considered.***
Whichever system is finally adopted will be interpreted as representing the conditions and skills to be valued in early education. However, the fact that we are more proficient in measuring some attributes than others should not suggest that we value those attributes more. This is especially apparent if such dimensions as language usage and general knowledge, which have long been the subject of measurement attention, are contrasted with social and emotional development, which have always been difficult to assess. Technical and value issues should not become confused. With this in mind, we propose that the assessment system should pursue reliable and valid measures of all dimensions of learning and development.

4. ***Validation studies must be an integral part of the assessment development effort.***
Because the state of the art in assessing young children is not far enough along to make it possible to create immediately the type of assessment system we envision, development efforts must be accompanied by on-going evaluation and validation studies. Although the final set of assessment instruments intended for administration to teachers, parents, and children must be short enough to meet time and cost constraints, during development more in-depth observations should be conducted to verify that conclusions reached on the basis of one-time assessments are valid representations of what children can do in natural classroom

settings. The congruence among the four data sources, parent reports, teacher reports, child profile and child portfolio, should also be evaluated to check for biases and to determine which data sources are the most valid and feasible means to use in assessing each of the five dimensions. Validation studies should be planned with special attention to the appropriateness, reliability, and validity of assessment procedures across diverse populations of children and families.

Issues Concerning Specific Components

Parent Report: Although parents can provide useful information on all dimensions of early school experience, their contribution may be greatest to our understanding of physical well-being and social and emotional development. Our ability to obtain reliable parental information about their children is currently limited technically and will require further development. In addition, our knowledge base about the relationship of home learning settings to early school experience is incomplete. The focus of this component should be descriptive rather than prescriptive, helping us to know more about the learning environment of the home as it relates to what children bring to school. Care must be taken to respect cultural differences that make our society so rich.

Teacher Report: Work has been done in this area of assessment, though not in the systematic, large-scale manner that is proposed. To reduce bias in teacher judgment, the checklist should contain questions about observable behaviors rather than global child characteristics. In addition, the behaviors should be presented in a non-normative manner. The use of the Teacher Report is most appropriate and best developed for providing information about language usage, cognitive functions, and general knowledge. With further exploration, it should also be applicable to social and emotional adjustment and approaches to learning.

Profile of Children's Skills, Knowledge, and Development: Because of the particular difficulties involved in gathering data from individual children, and because this is a source of information obtained at a single time point, it is vitally important to develop reliable and valid tools for this component. Substantial work has been done in assessment of individual children, but for purposes that differ from those of the Early Childhood Assessment. This component can provide valuable information on all dimensions of early learning and development. However, items relating to social and emotional development and approaches to learning will require more development than those concerning physical well-being, language usage, cognition, and general knowledge. Testing young children is labor-intensive; good information requires that the assessment be individually administered. With their responsibilities to the rest of the students in a given class, it may not be feasible for classroom teachers to administer this assessment. If not, the assessment should be administered by individuals who are well-trained and supervised.

Performance Portfolio: Portfolio assessment is a new strategy with very little research concerning its measurement properties. Although it appears to hold much promise, a great deal of work must be done before it can become a reliable component of the Early Childhood Assessment. It is anticipated that the Portfolio will contribute to our understanding of approaches to learning, cognition, social and emotional development, and general knowledge.

IMPLEMENTATION

Timely development of the assessment program is of great importance. However, a tension exists between using available instruments developed for other purposes or delaying the acquisition of information by developing new instruments matched to this particular objective. Although some instruments have been used previously for related purposes, reliable and valid instruments simply do not exist that will fulfill all of the functions we have proposed. Moreover, we are suggesting a system of assessment, not a set of separate indicators. To avoid redundancies of effort and cost, the Early Childhood Assessment will require development as a systematic network with closely interlocking parts. Because this is a different type of assessment endeavor from previous efforts, and because of our concern that the instruments used be valid for their intended purposes, we conclude that existing instruments can serve only as resources for the development of the Early Childhood Assessment. Substantial modification of these instruments will be necessary and new strategies and items must be generated.

Given the complexity and importance of this assessment program, what is a feasible timeline for research, development, and utilization of information? Our primary constraint is technical knowledge, which varies both along the dimensions of early childhood learning and development and the components with which we have proposed to assess them. The task of developing and validating new instruments appropriate for young children will be extremely time-consuming. Currently available measures are inadequate and are based largely on outmoded psychometric techniques of the 1920s. Nevertheless, we believe that the development of an integrated system of assessment is feasible as well as desirable. As this development process proceeds, we stress that the Early Childhood Assessment should not be considered complete until all components are developed, field tested, validated, and implemented.

Moreover, we suggest that data collection occur in three-year cycles. This recommendation is based on two considerations. First, it is assumed that change that might be tracked by the Early Childhood Assessment will occur at a slower pace than a yearly indicator program might show. From that perspective, it makes sense to assess less frequently. Second, the proposed system of assessment is necessarily complex, intensive, and as a result, very expensive. Quality of data is a worthwhile trade-off against frequency of data collection.

If we do not take time to develop an integrated and appropriate assessment system, several negative consequences may occur. First, quick development work is quite likely to produce weak or uninterpretable indicators. Unhappily, such indicators might be seized upon and used for want of anything better. Decisions based on such instruments could be harmful to children and their families. Second, past experience indicates that accurate assessment in early childhood is particularly challenging to achieve. We do not have measures that are sensitive enough to discern subtle differences in children. Thus, we run the risk of substituting poor short-term proxies for accurate long-term predictors. This was made quite plain in the history of Head Start evaluations when short-term cognitive testing suggested program failure although longer-term, more thorough assessments revealed positive "sleeper effects." Such premature assessment could be harmful not only to children, but to early childhood programs. Third, single source, non-objective ratings, such as teacher judgments about who should be retained in grade, have been shown to yield biased accounts of children's performance and abilities and to lead to inappropriate placements and inadequate programming for children. To stave off such inequities, we firmly underscore that any new assessment must obtain information about young children from multiple perspectives rather than relying on any single data element.

National Commission on Early Childhood Assessment

In the context of the development requirements just enumerated and the policy issues addressed earlier, we recommend the establishment of an independent, National Commission on Early Childhood Assessment. The purpose of the National Commission on Early Childhood Assessment would be to initiate and supervise a program of development and to oversee the implementation of the Early Childhood Assessment delineated in this report. The Commission would be established along the lines of other similar independent oversight groups created by the National Academy of Sciences and the National Academy of Education. It should meet regularly, and should participate in the organization of a coherent, focused program leading to the development of integrated assessments of children's physical well-being and motor development, social and emotional development, approaches to learning, use of language, and cognitive skills and general knowledge. It is critical that the Commission be established as a long-term effort. The tasks involved in developing and overseeing the implementation of the Early Childhood Assessment will not yield to anything other than a long-term commitment, and we believe that the Commission should exercise oversight throughout this process.

Finally, the Commission should have the responsibility of evaluating the assessment itself by providing ongoing information and feedback regarding the extent to which the assessment system is meeting its own objectives. In short, the Commission should review the results of the Early Childhood Assessment, determine the validity of the data collected, obtain feedback from participants in the assessments, and release information to the field and to policy makers.

CONCLUSION

The task of developing and implementing the Early Childhood Assessment provides us with a new set of opportunities and responsibilities. The system of assessment that we propose is a new venture and it is put forth with an appreciation for the complexity of child development. We have a vision of an assessment that addresses the whole child in an integrated manner. However, we cannot adequately stress the importance of the multidimensional approach as the system moves through development to implementation. The Commission on Early Childhood Assessment will be responsible, in part, for preserving this multiple perspective. If we attempt this task with less than the comprehensive approach outlined here, the data will be impoverished and the picture we construct from the data will be distorted. A complete picture requires consideration of all dimensions of the growing child from as many perspectives as possible.

CHAPTER 2

GOAL 2: HIGH SCHOOL COMPLETION

GOAL 2: **By the year 2000, the high school graduation rate will increase to at least 90 percent.**

Objectives:

- The nation must dramatically reduce its dropout rate, and seventy-five percent of those students who do drop out will successfully complete a high school degree or its equivalent.
- The gap in high school graduation rates between American students from minority backgrounds and their non-minority counterparts will be eliminated.

GOAL 2 RESOURCE GROUP ON HIGH SCHOOL COMPLETION

RESOURCE GROUP MEMBERSHIP

Convener

Edmond Gordon	Yale University, Connecticut

Members

Janet Baldwin	American Council on Education, General Education Development (GED), District of Columbia
Eve Bither	Maine State Department of Education
Jose Cardenas	The Intercultural Development Research Association, Texas
Noreen Lopez	Illinois State Department of Education
Steve Nielson	U.S. West Corporation, District of Columbia
Aaron Pallas	Michigan State University, School of Education
Rafael Valdivieso	Hispanic Policy Development Project, District of Columbia
Richard Wallace	Pittsburgh School District, Pennsylvania

GOAL 2 TECHNICAL PLANNING SUBGROUP ON HIGH SCHOOL COMPLETION

TECHNICAL PLANNING SUBGROUP MEMBERSHIP

Leader	
Rafael Valdivieso	Hispanic Policy Development Project, District of Columbia
Members	
Janet Baldwin	American Council on Education, General Education Development (GED) Testing Science, District of Columbia
Jomills Braddock	Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students, Maryland
Barbara Clements	Council of Chief State School Officers, District of Columbia
James McPartland	Johns Hopkins University, Center for the Social Organization of Schools, Maryland
Gary Natriello	Columbia University, Teachers College, New York
Aaron Pallas	Michigan State University, School of Education
Andrew Porter	University of Wisconsin, Center for Educational Research
Floraline Stevens	Los Angeles School District, California

September 4, 1991

TO: THE NATIONAL EDUCATION GOALS PANEL

FROM: EDMOND GORDON, CONVENER
THE RESOURCE GROUP ON HIGH SCHOOL COMPLETION

GOAL 2 RESOURCE GROUP

STATEMENT ON THE TECHNICAL PLANNING SUBGROUP'S REPORT

Summary of the Technical Report

Goal Two states, "By the year 2000, the high school graduation rate will increase to at least 90 percent." While the U.S. Bureau of the Census' Current Population Survey provides annual national data on high school completion rates, currently we do not collect accurate and comparable data at the state level. In particular:

- There is no agreement among states as to who is a graduate, who is a dropout, or even who is a student.

- Our current systems for counting graduates fail to distinguish adequately among the different ways of completing high school, including receipt of a regular high school diploma, another type of diploma, a certificate of completion or attendance, or a high school equivalency credential.

- Each year, more than 3 million school-aged children move between states, between counties within a state, or to or from locations abroad. Others move from private schools to public schools or vice-versa. The vast majority of public school districts lack the capacity to trace students across these boundaries. As a result, the enrollment status of many students is unknown.

- Under our current data collection systems, we cannot tell what proportion of students entering ninth grade in the fall of 1987, for example, actually completed high school by the spring of 1991, or will have completed high school by 1993.

- The most prominent of our current state systems for measuring high school completion, the "Wall Chart" graduation rate, does not adjust accurately for migration, retention in grade, or ungraded students. The resulting figure is widely recognized as inadequate.

The Planning Subgroup proposes a Voluntary State/Local Student Record System (VS/LSRS):

- To encourage states and districts to employ consistent, standardized definitions of a graduate, a student, and a dropout.
- To enable the states and nation to monitor accurately the awarding of different types of high school completion credentials.
- To help schools keep track of students who transfer to other schools, thus providing a more accurate accounting of where and when students are enrolled.
- To describe the experiences of cohorts of students as they move through school, and to end our reliance on flawed synthetic statistics.
- To make the Nation's Report Card less susceptible to apparent changes arising from the different ways we count students, dropouts, and graduates, and more responsive to real changes in patterns of high school completion.
- To provide complete and accurate information on high school completion in each state and the nation overall.

The Planning Subgroup recommends that the following steps be undertaken to develop the VS/LSRS:

- Compile current state practices and current plans for system development at the state and local level.
- Convene a national panel of experts on information system design.
- Convene a national task force on data elements.
- Develop prototype systems for the collection and maintenance of student records.
- Provide technical assistance to states, school districts, and local schools.

- Provide seed money to states.

The steps the Planning Subgroup proposes will have the following short-term effects:

- As early as 1993 or 1994, the data on high school completion collected by schools, districts, and states will be more standardized and comparable.
- As a result, the data will be more accurate, and the overall quality of our information on progress toward Goal Two will be greatly improved.
- Over time, our data collections will be successively more complete and accurate.
- Because the proposed system is expandable, it will provide ever more detailed information on progress toward Goal Two and the other National Education Goals.

Position on the Technical Report

We accept and endorse enthusiastically and wholeheartedly the report prepared by the Goal Two Technical Planning Subgroup. The report reflects thoughtful weighing of a number of considerations and issues inherent in our original suggestions for a national reporting system.

We want to make two additional points. First, close attention should be paid to the interrelating of the Voluntary State/Local Student Record System (VS/LSRS) and the data collection systems beyond the secondary school level. We are referring to data collection for the General Educational Development (GED) program, administered by the American Council on Education (ACE) and the Adult and Higher Education Systems, which is not normally under the jurisdiction of the Chief State School Officer. We must make sure that persons who complete the equivalent of a secondary education in these systems are also counted.

One of the many appealing aspects of this plan is that it builds on the current state of affairs among the states. Many states have been working towards standard definitions for dropouts and school completers. While only five or so states now have state-wide collection systems, many states are interested in moving in that direction. If national leadership is provided as proposed, along with a set of incentives sufficient to motivate participation in the VS/LSRS, we think it will be only a few short years before we have full participation by virtually all states.

THE GOAL 2 TECHNICAL PLANNING SUBGROUP REPORT ON HIGH SCHOOL COMPLETION

Submitted to

The National Education Goals Panel

through

THE GOAL 2 RESOURCE GROUP ON HIGH SCHOOL COMPLETION

for presentation at

The National Education Goals Panel Meeting

September 4, 1991

THE GOAL 2 TECHNICAL PLANNING SUBGROUP REPORT ON HIGH SCHOOL COMPLETION

Introduction

This document represents an initial statement by the Technical Planning Subgroup of the Resource Group for Goal 2, High School Completion. In this report, we consider the technical issues raised by the Resource Group's Interim Report. We begin by re-examining the data needs presented by the charge to measure progress toward Goal Two. We then describe a voluntary state/local student record system designed to produce such data, and consider several implementation issues, including timing and data quality. We conclude by placing the system in the context of a more comprehensive effort to measure progress toward all of the goals simultaneously.

We are grateful to the Resource Group for their thoughtful interim report proposing both short-term and long-term strategies for measuring progress toward Goal Two. As with the Resource Group before us, we reaffirm the importance of Goal Two to the well-being of our children and youth and our schools. While a high school diploma or its equivalent may tell us little about what an individual has learned, *not* having such a credential places severe limits on educational and occupational opportunities. Because completing high school is an important determinant of adult life chances, the Nation must have an accurate accounting of the proportion of young people who achieve a high school diploma or its equivalent. The high school completion rate enables us to gauge the success of individuals and of the education system overall.

The Problem

It may seem surprising that the development of a high school completion rate would require technical planning. In the abstract, the problem seems quite simple: we merely enumerate the high school-aged population, and determine what proportion of these youth have completed high school. But the apparent ease of this task masks some long-standing problems in the measurement of high school completion and dropout rates. *In fact, currently we do not collect accurate and comparable data at the State level.*

A lack of consistent definitions. There is no agreement across States as to who is a graduate, who is a dropout, or even who is a student. For example, some States include high school equivalency recipients with regular school diploma recipients, while others report only regular high school diplomas. Similarly, some States award a regular high school diploma to special education students, while others award a certificate of completion or attendance.

Consequently, it is impossible to compare data across States, or even to use an aggregation of the State data to describe the Nation as a whole.

Failure to distinguish among alternative credentials. A related problem is the inability of current data collections to distinguish among different ways of completing high school. There are at least four types of high school completion credentials: receipt of a regular high school diploma, another type of diploma (e.g., a Regents diploma), a certificate of completion or attendance, and a high school equivalency credential. Because educational policies and reform strategies may influence the distribution of these credentials, and because they may represent differing levels of educational accomplishment, it is imperative that our data collections distinguish among these different completion categories.

Failure to trace students who move. Each year, more than 3 million school-aged children move between states, between counties within a state, or out of the country. Others move from private schools to public schools, and vice versa. But under the current system, schools are ill-equipped to determine whether a student who leaves one school has enrolled in another approved educational program. The inability to distinguish between transfers and dropouts muddies the calculation of high school completion rates.

Inability to trace individual students through their school careers. When we think of a high school completion rate, typically we are asking the implicit question, "Of the students who entered ninth (or some other) grade in a particular year, what proportion completed high school within four (or some other number of) years?" Answering this question requires data on a *cohort* of students who experienced some common event (i.e., entering a particular grade at the same time). Studying a cohort is like taking a moving picture of the same group of students over time. But our current data collections consist of taking snapshots at different points of time. The problem is that the students represented in one snapshot may not appear in the next. Some of this year's ninth graders will need three years to complete high school, while others may need four, five or more. Any "snapshot" of graduates taken in the future will fail to capture all of the eventual graduates from this year's ninth grade class. As a result, the snapshots which comprise our current data collections introduce serious errors into our estimates of the high school completion rate.

Although these problems with the available data are well-known, they have been ignored in most reporting on high school completion rates. For example, the data on high school completion reported in past years in the Secretary of Education's "Wall Chart" are subject to each of the difficulties noted above. *Those data are not good enough.* The National Education Goals Panel's mandate to assess progress toward the six National Education Goals affords an unprecedented window of opportunity to use information strategically to promote student achievement. Many States have already made investments in their data collection systems, and others are about to do so. The system we are proposing will capitalize on these investments, and help to generate momentum for further investment.

Data Needs

Although there is some overlap in the membership of the Resource Group and the Technical Planning Subgroup, both groups brought distinctive backgrounds, experiences, and perspectives to bear on the political, social, and technical issues involved in measuring progress toward Goal Two. In our deliberations, we felt free to criticize, modify or embellish the recommendations contained in the Resource Group's Interim Report when it seemed appropriate to do so. We believed that extending the conversations about these issues to an ever-larger audience would eventually lead to a more comprehensive and unified approach to a set of problems that have been resistant to simple solutions.

We began our deliberations by considering the types of data needed to measure progress toward Goal Two, increasing the high school completion rate. Among the direct data needs, we counted high school completion rates for various groups defined by race/ethnicity and gender, and for youth of varying ages. This information is needed for each State and for the Nation overall. Such data should include separate counts for the major ways of classifying high school completers (i.e., regular high school diploma, another type of diploma, a certificate of completion or attendance, or a high school equivalency credential).

The inclusion of gender is an extension of the recommendations of the Resource Group. This recommendation is included because not only do school completion rates differ for different racial/ethnic minority groups, but there also are differences between the sexes in school completion rates *within* racial/ethnic groups. The simple reporting of high school completion rates of males and females and of differing racial/ethnic groups conceivably could mask important patterns of progress toward Goal Two.

In our deliberations, we also identified several other types of data we believed could be useful in measuring the Nation's progress toward Goal Two, and that might be able to contribute to that progress. We felt that our efforts to improve the high school completion rate could be better focused with regular reporting of how school completion and dropout rates vary by the type of school attended, and of how the social background characteristics and school experiences of high school completers differ from those of their peers who dropped out of school. We believe, for example, there is much to be learned from comparing the grade retention histories of high school graduates and non-completers.

This example is merely illustrative, however, and we wanted to fashion a data collection and reporting system that could support a wide range of queries, not all of which could be anticipated in advance. Ten years ago, it might not have occurred to a national audience that retention in grade might plausibly be linked to the likelihood of school completion. Because the most pressing educational issues ebb and flow over time, we judged that a flexible structure capable of self-correction and refinement would serve the Nation best.

Moreover, we were taken by the Resource Group's call for a data system that would serve the needs not only of our Nation's political leaders, but also would place information in the hands

of those who do the bulk of the work of educating American children and youth -- teachers and other building-level staff. As the Resource Group noted, school-based information systems, by providing relevant information on students' backgrounds, performance and behavior, have the potential to enable school personnel to make better decisions about teaching and learning, and thus improve the educational prospects of children and youth -- and increase their likelihood of completing high school.

Recommended Strategies

Voluntary State/Local Student Record System

The core of our proposal is a voluntary state/local student record system (VS/LSRS). Such a system would store longitudinal information on individual students at the school building, district or State level, and would allow summaries or abstracts of those data describing groups of students to be forwarded to the State level, and, ultimately, to the National level. The key to this system is the *standardization of data elements*. The development of common definitions, rules and standards to be used in the recording, maintenance and storage of the relevant data is critical to the effort. By standardizing the data contained in the VS/LSRS, we can be assured that the data reported by one school are comparable to those from a neighboring school, or that one district's high school graduate count is based on the same definition of a graduate as the next district. By maintaining consistency in the collection and reporting of data at the local level, we can assure the comparability of data, thereby increasing the quality of data in each State and for the Nation as a whole.

The system we are proposing addresses each of the problems we discussed earlier. First, as we note above, this system encourages States and districts to employ consistent, standardized definitions of who is a graduate, who is a student, and who is a dropout. The data gathered by each State and aggregated to the national level thus will be comparable across States and over time. The use of such definitions will guard against changes in reported high school completion rates that are solely attributable to changes in definitions or accounting methods.

Second, the system will enable the States and the Nation to monitor accurately the awarding of different types of high school completion credentials. The current data collections do not always maintain potentially relevant distinctions among different types of credentials, sometimes equating certificates of attendance , regular high school diplomas, and high school equivalency credentials. Tabulating these various credentials separately allows maximal flexibility and detail in constructing high school completion rates.

Third, the system will help schools keep track of students who transfer to another school, thus resulting in a more accurate accounting of where and when students are enrolled in school. By routinizing the exchange of standardized school record information, the system will

encourage schools enrolling new students to request student records from the sending schools, thus notifying the sending schools of the current enrollment status of students transferring to other schools. By helping to reduce errors in enrollment status, the system will promote more accurate school completion and dropout statistics, since a major problem in calculating such rates is determining whether a given student is enrolled.

Fourth, the system will describe the experience of *cohorts* of students as they move through school, and end our reliance on flawed synthetic statistics. We can get the most accurate accounting of who is completing high school by tracing individual students over time. *In sum, the system we are proposing will provide complete and accurate information on high school completion in each State and the Nation overall.* These changes will make the Nation's Report Card less susceptible to changes that result from the way we count students, dropouts and graduates, and more responsive to real changes in patterns of high school completion in our Nation's schools.

A Modular System

We envision a Voluntary State/Local Student Record System that is modular, with a common core of biographical information and the capacity to link that common core to other data modules describing other aspects of school experiences and academic performance. At a minimum, we believe that the student biographical information should include basic identifying information, sex, race/ethnicity, date of birth, place of birth, and enrollment histories. These data elements should allow for the production of the data needed to measure progress toward Goal Two. We recognize, however, the desirability of generating a national consensus on what information should reside in the VS/LSRS, a strategy for which is discussed later.

In addition to the common core of information contained in the VS/LSRS, the system is designed to allow local schools and districts to supplement the core with other information deemed relevant. For example, a module on school experiences might include data on courses taken, and participation in special programs such as Chapter I, special education, or gifted and talented programs. Similarly, an academic performance module might include evaluations of student performance such as grades and test scores. Some schools and districts with students participating in state or local programs might wish to record program participation in the system. Others may wish to record transportation information, or disciplinary actions, or social service referrals. As before, these elements are by way of example. They are intended to be neither exhaustive nor prescriptive. We envision design protocols for many of these options.

We recognize the varying capacities and resources available to help States to implement the VS/LSRS. Some States already have much of what we are proposing in place; others are still dependent on manual data collection systems that are quite primitive by current standards. The various States thus will find themselves at various stages in the development of the

VS/LSRS. The eventual success of the system is dependent on gaining the participation of all the States in the maintenance and reporting of the core biographical information. Hence, this must be the first priority of system implementation. At the same time, we see the potential for significant advances in the information processing capacities of schools and school districts through supplementing this core information with information on school experiences and academic performance. At the conclusion of this report, we discuss possible connections between the measurement of progress toward Goal Two and the measurement of progress toward the other National Education Goals.

A key advantage of the VS/LSRS we propose is that it builds on existing systems. We felt it important to recommend an approach that would minimize the duplication of ongoing efforts at the State and local levels to improve the processing of student information. There are, for example, pilot studies in progress attempting to transfer transcripts and school records electronically, using common formats developed by the National Center for Education Statistics in conjunction with a number of other organizations, including the Council of Chief State School Officers. What we are proposing is a logical extension of these existing efforts.

Who it Benefits, and Why

The system we are proposing can improve the information-processing capacity of the Nation, the States, school districts, and individual schools. Since information is such a valuable resource in making decisions about schools and schooling, the development of the proposed VS/LSRS has the potential to improve the quality of our efforts to reform and restructure American education.

Federal and state levels

We need valid and comparable information to produce comparable state data and national data. At the federal and state level, the proposed student record system will provide national, state and local estimates of the number and proportion of students who leave school each year prior to receiving a high school diploma or its equivalent. These estimates will also be reported for subgroups defined by student gender and race/ethnicity. Thus, the system will allow the Nation to track the dropout rates of, for example, black males or Native American females, and to judge whether these rates are changing over time, and whether they differ from the rates for non-minority youth. The system also will provide parallel information on youth completing high school, distinguishing among several alternatives, such as a regular high school diploma, another type of diploma, a certificate of completion or attendance, or a high school equivalency credential.

Local level

At the local level, the data produced by the VS/LSRS are not oriented primarily toward comparisons across schools and districts. Rather, the key is that such data will be *useful* to

staff in schools and districts as they carry out their mission to educate children and youth. At the district and school level, the proposed student record system will facilitate transfer of information across school boundaries, and thus lead to improved accounting of school dropouts and graduates. This system can also yield information on the characteristics of students who are at risk of school failure and/or dropping out, and thus provide a guide to the allocation of resources where they can do the most good.

Additionally, a fully implemented VS/LSRS can serve as an early warning system, providing information on students' performance and experiences that signal the likelihood of leaving school prior to graduation. By using the system to identify potential school-leavers, district and school staff can target the available resources to those students who are most at risk of dropping out of school, and perhaps stimulate them to stay in school. In this way, the VS/LSRS, a tool for measuring progress toward Goal Two, can help to promote progress toward the Goal.

Implementation strategies

Our goal is to capitalize on the existing, ongoing work of the States in refining their student information systems, particularly their data on school dropouts and graduates. We wish to minimize the duplication of efforts in the development of such systems. We believe the federal government, in cooperation with other interested groups, can provide leadership to the States by funding the development of prototype information systems and extending technical support to the States in their implementation.

The system we envision is truly voluntary. No one can mandate State participation, and individual States will have differing levels of interest and varying levels of resources available to implement the system we envision. Already there is considerable variability across States in their existing information processing capacity; this variability may persist for some years to come. We expect, however, that the historic tripartite commitment of the President, the Governors, and the Congress will create a climate in which conforming to the consensual agenda will be strongly encouraged, both across States and within them. As we look several years into the future we hope and expect that *all* of the States will be active participants in this new effort to progress, and measure progress, toward the National Education Goals. Only with the complete participation of all States, and all schools within them, can we ensure high quality data at the local, State and national level.

Concrete steps

We recommend that the following steps be undertaken to develop the VS/LSRS:

1. **Compile current state practices and current plans for system development at the state and local level.** The development of information systems and data elements should build on existing and proposed systems. Currently there is no compendium of the range of student information systems in use in elementary and secondary schools

and school districts in the U.S. Our discussions of how to proceed need to be informed by our recent history of successes and failures in this realm.

2. Convene a national panel of experts on information system design. We recommend that the National Center for Education Statistics convene a national taskforce to examine the technical issues involved in the design and implementation of local school-based information systems. The members might be drawn from states and school districts that have well-developed information systems already in place or currently under development, and should include a diverse set of stakeholders, including teachers and administrators. This group could recommend strategies for system design and development to NCES and other groups.

3. Convene a national taskforce on data elements. Even more challenging than the design of a generic information system is building consensus on the data elements that will populate the system. Already NCES and the Council of Chief State School Officers have devoted a great deal of energy to the standardization of data elements in NCES' routine data collections. These two groups should take the lead in considering two questions: what data elements should comprise a common core of information in student record systems, and how should these elements be measured? The first of these questions obviously involves a wide range of stakeholders in the education system, including building-level staff such as administrators, teachers and counselors, as well as parents and students. All of these groups will need to be adequately represented in the national debate over the standardization of student records.

4. Develop prototype systems for the collection and maintenance of student records. NCES should take the lead in designing generic student record systems that can be adopted by states, districts, and schools. We recommend that NCES compete a design contract for the development of a generic student record system that maintains data at the local level but provides a mechanism for aggregating information to the State level. The product of this contract should be a set of tools, specifications, and manuals designed to assist states and local school districts to implement the VS/LSRS.

The system designs will vary according to the extent to which the proposed systems extend beyond the information needed to measure progress toward Goal Two. A system that integrates the data needed to measure all of the National Education Goals, for example, would be considerably more complex than a system that is less comprehensive in scope. The prototype designs should take into account several options for integrating various data modules with the core school enrollment information.

5. Provide technical assistance to States and school districts. Technical assistance to States should include training State education agency staff to provide assistance to schools and districts within the State. Such technical assistance will include manuals

and hands-on training, as well as assistance in hardware and software selection. We also recommend that the training incorporate skills in data collection and analysis, to allow States to more fully exploit the available data to address educational policy concerns.

6. Provide seed money to States. We recommend that NCES provide resources to requesting States to aid in the development of the VS/LSRS. Since States are in different stages of development, they will likely use such seed money for different purposes, but virtually every State can benefit from further development work. We further recommend that the budget of NCES be augmented to include the costs of the seed money and technical assistance activities we have described. Alternatively, the National Forum on Education Statistics might take the lead role. Ultimately, however, it is the responsibility of States and the schools and districts within them to implement the VS/LSRS. The system is, by definition, a voluntary one. Moreover, the costs of implementing such a system will largely be borne by States and districts. Our hope is, however, by providing appropriate technical assistance and seed money, and by demonstrating the manifold payoffs to such a system, that States and districts will be willing to invest in their information-processing capacities.

Short-term effects on data quality by '93 or '94

We see the process of improving the quality of data on school completion as iterative, with successive refinements at the level of individual schools, districts and States. Each time an individual school, district or State collects student information using consistent, agreed-upon procedures and definitions, the various data collections in the Nation become more comparable, and the overall quality of U.S. data increases. Thus, we can expect to see substantial improvements in the quality of data within the next three years. Even though it may take a decade for all of the schools, districts and States in the Nation to adopt and implement consistent, standardized data collection procedures, there are many successive improvements to be gained along the way.

Alternatives we ruled out

We considered a number of alternatives to the system that we have proposed. Each was determined unworkable, either on political, fiscal, or technical grounds (or some combination). We describe below three of the more prominent alternatives we considered.

National student record system

One alternative to our preferred system is a national student record system that maintains detailed, student-level information on all students attending U.S. public schools. The advantage of such a system is that it provides the most flexible data structure for monitoring educational performance, including measuring progress toward Goal Two and the other

National Education Goals. Such detailed information could be aggregated to the school, district, State, or National level, depending on the nature of the specific inquiry.

There are several key problems with such a system, however. First, the centralization of such a mass of detailed information on individual students raises the specter of Big Brother. The Congress and the States have enacted much legislation designed to protect the privacy rights of students and their families. A large, centralized system may threaten that commitment, as both technical and sociopolitical concerns over data access and security may be difficult to resolve.

Second, the development of centralized computer systems runs counter to an apparent trend is more toward the development of decentralized or distributed systems. However, some States, such as Texas, have developed centralized systems, so clearly it is possible to do so at the State level. But even the largest States comprise only a small fraction of the Nation's population of children and youth. Thus centralized State systems provide little guidance as to the technical feasibility of a system of much larger scale. More important than even technical capacity, however, is ease of access and utility. Sprawling centralized systems tend to be more unwieldy and less responsive to local needs.

Third, even if we were able to implement such a system, the cost might be prohibitive. For each of the more than 80,000 public elementary and secondary schools to have access to the system, there would need to be an enormous computer network. This type of system would require a substantial outlay of money at the outset, in order to bring schools "on line". Since the system would be unable to produce State data unless all of the schools in the State were connected to the system, there is less margin for the phasing in of the system and the staggering or deferral of costs.

Samples of schools and students

We also considered a data collection system that would sample schools and/or students at the National level and within each State. Such a system had three distinct disadvantages. First, as with all sample surveys, the data produced would be estimates, not population values. Generating reliable estimates at the State level might require very large numbers of students and/or schools within States. Consequently, there may be little gain from sampling. Second, a sample survey would not provide data on every school. As a result, only those schools sampled in the survey would be able to compare themselves with the data produced by each State or the Nation overall. Most schools would not be able to judge their progress toward the Goal. Third, a federal sample survey undermines the notion of shared responsibility that undergirds the National Education Goals effort. Making progress toward the Goals is not merely a federal responsibility; rather, it is a responsibility that spans the executive branch, the Congress, the States, and school districts and schools. A data collection system that involves schools and school districts in the effort to measure progress toward the National Education Goals may produce a stronger commitment to those goals.

Narrow, completion-focused data collection

We also ruled out a narrow data collection focused exclusively on the collection of high school completion data. Such a system might have been what we were expected to recommend. While this kind of narrow data collection could improve the quality of our estimates of high school graduation rates in much the same way as we have recommended, it simply would not tell us much. There is still a great deal we do not know about the determinants of high school completion, especially the impact of current and recent educational reform proposals on high school completion rates. A narrowly defined system would preclude asking many of these questions. This is why we strongly recommend a system that has the capacity to link high school completion data to other relevant information on students' backgrounds and school experiences. Not only can a more comprehensive system assess progress toward several goals simultaneously, but there are substantial economies of scale that stem from building systems that are integrated and do not duplicate data collection, storage and retrieval.

Timelines

Task	Date
1. Compile current state practices and current plans for system development at the state and local level	by middle of 1992
2. Convene a national panel of experts on information system design	by middle of 1992
3. Convene a national taskforce on data elements	by middle of 1992
4. Develop prototype systems for the collection and maintenance of student records	1993
5. Provide technical assistance to States and school districts	1994
6. Provide seed money to States	1994

Limitations/Special Issues

No system is free of limitations or special issues that can place boundaries on its implementation and utility. We note below some of the most important limitations and special concerns. Doubtless others will be identified as these recommendations are debated and discussed in a wider forum.

Private schools and adult education providers

Any system designed to assess progress toward the National Education Goals must take account of those students who do not attend public elementary and secondary schools. Approximately one student in eight is enrolled in a private school, and others are enrolled in adult education programs outside the traditional K–12 system. It simply does not make sense to exclude such schools and programs, and the children and youth who attend them, from our measures of progress. In addition, the maintenance of comparable student records in these schools and programs would facilitate the exchange of data with the public schools. This would enable both types of schools to better monitor student movement between these schools and maintain more accurate counts of high school completers and dropouts. In keeping with the spirit of the voluntary student record system that we have described, we hope that independent and other private schools and adult education providers will choose to participate. We recommend that the same types of technical assistance be extended to these institutions as to public schools.

Similarly, some children and youth are institutionalized during their school–going years. While we are concerned about the educational opportunities of America's institutionalized youth, we recognize that institutions serving children and youth may lack the capacity and resources to maintain the kinds of records that are so common in schools. Careful thought must be given to how best to incorporate these institutions into our data systems.

Providing appropriate information to teachers

The most significant advantage of the VS/LSRS we have outlined is that it will produce information that is useful at the local site level. Of course, some of the resistance to comprehensive student information systems stems from the fear that information in those systems might be used in ways that are not in the best interests of children and youth. For example, there is a persistent concern that the labels that are applied to children in school records can follow and limit them for years to come, even if those labels were inappropriate or inaccurate. Children labeled as mentally retarded or learning disabled often are treated differently by school staff than similar children who are not so labeled.

There is no doubt that some school staff members use labels inappropriately to summarize and stereotype children, and that these labels, rather than children's actual educational needs, determine the access to and quality of the educational services they receive. The solution to this problem is not, however, to throw out useful information about students' academic performance and behavior or backgrounds. Rather, we believe that this problem can best be handled by educating school staff about the meaning of labels, and training them in the proper use of school record information. Most school staff do not set out to falsely stereotype children receiving special services; nor do they wish to thwart their educational opportunities or progress.

Some teachers pride themselves on not knowing anything about a child before she or he arrives in class. They claim that such ignorance protects them from labeling children and treating them inappropriately. In fact, however, labeling is most likely to be inappropriate when labels are based on limited information. Broadening the base of information makes inappropriate labeling less likely. If teachers are to be treated as professionals, as is their wont, they must be expected to exercise independent judgment in the diagnosis of educational problems. No one would expect a physician to begin a delicate surgical operation without first having adduced a patient's medical history. It is equally fatuous to expect a teacher to teach children without an understanding of their current conditions, previous performance, and backgrounds.

Local quality control

A data system is only as good as the data that are housed within it. To this point we have had little to say about control over data quality. We are most concerned with the quality of data entered into the system at the local level, since those data will be sent on and aggregated to produce State and National figures. There are no magic bullets with regard to data quality. But the VS/LSRS we have proposed does, we believe, have at least two features which may promote high quality data. First, the data reside at the local level, close to where the process of education is carried out. Errors in data, for instance, typically are more likely to be spotted and are more easily corrected at the local level than at, say, the State level. Second, the system is designed to produce data that are useful, thus providing school staff with an incentive to generate accurate information. When information is disconnected from the process of education, there is little reason to expect school staff to invest in the production of high quality data. Our hope is that the data contained in the VS/LSRS will be sufficiently useful to sustain the collection and reporting of high quality information on school completion and dropping out, as well as the other types of information we have discussed.

Privacy

A related issue concerns the privacy of the educational records of children and youth. In the early 1970s, parents and student advocates became increasingly concerned over the nature and accuracy of information contained in student records. This concern was fueled in part over issues regarding government collection of information, and specifically regarding information from special education assessments of children that resulted in detrimental labeling of students. The Freedom of Information Act of 1974 and the Family Educational Rights and Privacy Act of 1974 (known as FERPA or the Buckley Amendment) restricted third party access to educational records, including access by school personnel, on a "need to know" basis.

In light of these protections, care must be taken to ensure that student records are used in ways that are consistent with the current interpretations of the law. This may involve the removal of individually identifiable information from data forwarded from the school to the district, or from the district to the state. We envision relatively few circumstances when such individually identifiable data would be essential at the state level, however.

Potential of system to produce school-level data

We recognize the potential of this system to produce detailed comparisons of schools and districts. This is an "intermediate" type of information that lies between the more pressing needs for data at the National and State level, on the one hand, and the needs for data *within* schools and districts in the ordinary course of the practice of educating children and youth. It is conceivable that gathering and reporting data on individual school buildings and districts *could* be appropriate (i.e., comparing a particular school's dropout rate to itself over time), but we caution against the dangers inherent in inappropriate comparisons (i.e., comparing the rates of schools or districts that are fundamentally different in ways that might affect their dropout or completion rates).

The risk of inappropriate comparisons is likely to be minimized, however, in a system that makes it possible to put school-level data in the appropriate context. By going beyond simple accounting of dropout and completion rates, the VS/LSRS provides the tools for interested parties to make intelligent comparisons that take into account the ways in which schools and districts vary along several dimensions simultaneously.

Connections to Other Goals

The development of the VS/LSRS is closely linked to attempts to measure progress toward the other National Education Goals. In particular, we believe that the measures used to assess progress toward Goal One and Goal Three (i.e., the new American Achievement Tests) may well become part of the VS/LSRS. Information on health gathered at the time of school entry may become part of a child's permanent record, and may allow the Nation to gauge progress in preparing children for learning, as well as to assess the educational consequences of indicators of school readiness. Similarly, the assessments used to tap student achievement should become an integral part of a child's permanent record, providing information to building-level staff and beyond on what individual children and groups of children have learned.

In a full-blown implementation, the proposed VS/LSRS could embed school-, district-, and state-level information on high school dropout and completion in a more comprehensive record system. Such a system could report indicators of the health of the U.S. education system, as well as become a vehicle for more basic research on the performance of U.S. schools. Properly designed, this system could take account of the interrelatedness of the National Education Goals, and enable the Nation and the States to monitor progress toward all of the Goals simultaneously. The advantage of such an arrangement is that it allows us to assess whether policies designed to promote progress toward one Goal might have unanticipated consequences for progress toward one or more of the other Goals. We urge that the strengths and weaknesses of such an integrated system be a part of the debate regarding long-term strategies for measuring progress toward the National Education Goals.

CHAPTER 3

GOAL 3: STUDENT ACHIEVEMENT AND CITIZENSHIP

GOAL 3: **By the year 2000, American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter, including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy.**

Objectives:

- The academic performance of elementary and secondary students will increase significantly in every quartile, and the distribution of minority students in each level will more closely reflect the student population as a whole.
- The percentage of students who demonstrate the ability to reason, solve problems, apply knowledge, and write and communicate effectively will increase substantially.
- All students will be involved in activities that promote and demonstrate good citizenship, community service, and personal responsibility.
- The percentage of students who are competent in more than one language will substantially increase.
- All students will be knowledgeable about the diverse cultural heritage of this nation and about the world community.

GOAL 3 RESOURCE GROUP ON STUDENT ACHIEVEMENT AND CITIZENSHIP

RESOURCE GROUP MEMBERSHIP

Convener	
Lauren Resnick	University of Pittsburgh, Learning Research and Development Center, Pennsylvania
Members	
Gordon Ambach	Council of Chief State School Officers, District of Columbia
Chester E. Finn, Jr.	Vanderbilt University, Educational Excellence Network, District of Columbia
Asa Hilliard	Georgia State University
David Hornbeck	Education Advisor, Maryland
Richard P. Mills	Vermont State Department of Education
Thomas W. Paysant	San Diego School District, California
Claire Pelton	San Jose Unified School District, California
Terry K. Peterson	South Carolina Business Education Committee
Marshall S. Smith	Stanford University, School of Education, California

GOAL 3 TECHNICAL PLANNING SUBGROUPS ON THE NATIONAL ASSESSMENT SYSTEM[1]

TECHNICAL PLANNING SUBGROUP MEMBERSHIP

Leader	
Marshall S. Smith	Stanford University, School of Education, California
Members	
Eva Baker	University of California at Los Angeles
Chester E. Finn, Jr.	Vanderbilt University, Educational Excellence Network, District of Columbia
Edmond Gordon	Yale University, Connecticut
Silvia Johnson	Howard University, District of Columbia
Robert L. Linn	University of Colorado
George Madaus	Boston College, Massachusetts
Lauren Resnick	University of Pittsburgh, Learning Research and Development Center, Pennsylvania

[1]In addition to this Technical Planning Subgroup, two other Subgroups were created during 1991: Subgroup on Citizenship (Chair, David Hornbeck); Subgroup on NAEP (Chair, Richard Mills). These groups will be issuing their reports in early 1992. The preliminary work of these Subgroups may be found in Appendix A.

September 4, 1991

TO: THE NATIONAL EDUCATION GOALS PANEL

FROM: LAUREN RESNICK, CONVENER
THE RESOURCE GROUP ON STUDENT ACHIEVEMENT

GOAL 3 RESOURCE GROUP

STATEMENT ON THE TECHNICAL PLANNING SUBGROUP'S REPORT ON THE NATIONAL ASSESSMENT SYSTEM

Based on discussions of the Technical Planning Subgroup, Robert Linn prepared a paper that considered two broad issues associated with the Group 3 proposal for a nationwide assessment system aimed at improving achievement in American schools: (1) technical issues involved in a national assessment system in which several different exams (serving clusters of states) would be calibrated to a single national standard; and (2) standards of test validity associated with different uses of the proposed new assessment system.

Members of Resource Group 3 were invited to comment on the Linn paper. Replies were received from seven members of the group. Because of summer schedules, it was not possible to contact all members of the group. The following notes summarize the views of those who responded, clarified by further discussion with Linn and other psychometricians.

With respect to issue one, methods of calibrating cluster exams to a national standard, Linn outlined three broad options for calibration. Each option was analyzed in terms of technical problems that would need to be addressed if it were to be used in the nationwide system. Some members of the Resource Group wondered initially whether this implied that a multiple exam system might be impossible in practice. However, further discussion with Linn and consultation with other technical experts, suggested that none of the problems identified by Linn was insurmountable.

The Resource Group thus continues to recommend that a nationwide system including several examinations in each subject matter, all calibrated to a single national standard (which would also be shared by the NAEP monitoring test), be developed. This is thought to be the best means of accommodating the American tradition of state responsibility for education while at the same time promoting higher achievement and a shared standard throughout the nation. At the same

time, members of the Resource Group note that efforts to develop "cluster exams" will also yield a potential single exam for use throughout the nation should a calibration approach prove too unwieldy in practice.

With respect to issue two, validity standards, Linn's paper noted that standards of test validity depend on the intended use of the test and went on to discuss the kinds of evidence of validity that would need to be gathered before exams developed for the new nationwide system could be used for various forms of accountability. The broad impact of his argument, which reflects the standards of validity current in the field of psychometrics today, is that it may take some years after an exam is put into place for "highstakes" uses (e.g., granting of diplomas to individual students; provision of financial or other rewards and penalties for educators) to become appropriate.

However, validity studies appropriate to these uses can, for the most part, be conducted only after the exams are in place. The Resource Group therefore recommends that new exam development and deployment move forward accompanied by planning for validity studies that will permit high stakes uses as early in the decade as possible. We note further that it is particularly appropriate to begin with fourth grade exams, which will not dramatically affect students' life chances and which will therefore not require as extensive validity checks as high school exams.

THE GOAL 3 TECHNICAL PLANNING SUBGROUP REPORT ON THE NATIONAL ASSESSMENT SYSTEM

Submitted to

The National Education Goals Panel

through

THE GOAL 3 RESOURCE GROUP ON STUDENT ACHIEVEMENT AND CITIZENSHIP

for presentation at

The National Education Goals Panel Meeting

September 4, 1991

THE GOAL 3 TECHNICAL PLANNING SUBGROUP REPORT ON THE NATIONAL ASSESSMENT SYSTEM

INTRODUCTION

Technical Considerations in the Proposed Nationwide Assessment System for the National Education Goals Panel.

> National Education Goal 3: Student Achievement and Citizenship. "By the year 2000, American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter including English, mathematics, science, history and geography; and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment."

In 1990 the President and the Nation's Governors adopted a set of six National Education Goals. The National Education Goals Panel, composed of six Governors, four members of the President's Administration, and four members of the United States Congress, was created in July, 1990 to monitor and report on the Nation's and of States' progress towards meeting those goals. During the past year the Panel has worked on plans for both short-term reports, including preparations for release of its first annual report in September, 1991, and long-term plans for the development and implementation of more comprehensive mechanisms for monitoring progress in future years.

In carrying out its work, the Panel formed six Resource Groups to advise on the best indicators for monitoring progress towards the six goals. Interim reports from the Resource Groups were provided to the Panel in March 1991. These interim reports, together with a series of issues and questions concerning each goal, provided background materials for Panel hearings conducted in the spring of 1991.

In addition to providing recommendations for the short-term monitoring beginning in 1991, the Resource Group for Goal 3 recommended the creation of an "end-of-decade nationwide assessment system". The proposed nationwide assessment system is intended not only to provide a means of monitoring progress, but to contribute to the major changes in the nation's education system that are needed to achieve Goal 3.

The proposed nationwide assessment system is based on a vision of assessment that is closely linked to learning and instruction. Assessments would be based on clearly defined curriculum frameworks. The end-of-decade system would be radically different than testing programs that are currently in use in states and local districts throughout the country. The system would be based on "world-class standards" and use authentic performance tasks that would

require students to demonstrate higher-order reasoning and problem solving skills in each of the five content domains of Goal 3. In addition to monitoring progress in relation to Goal 3, the assessments are expected to provide a catalyst for educational improvements that will help assure that students and teachers put forth the effort that is needed for all students to achieve the high standards that will be established in each of the five content domains of Goal 3.

The proposed system is certainly ambitious. Indeed, it is unprecedented. The closest analogies in this country are the College Board's Advanced Placement (AP) program and parts of the Regents' Examinations in New York State. The tests in those programs are each based on a clearly specified curriculum and involve a combination of testing modes ranging from multiple-choice questions to more extended performance in response to open-ended problems. A major difference is that the AP and many Regents' courses are designed for an education elite and taken by only a relatively small fraction of college-bound students. It is an open question whether a single examination could well serve high school students from vocational and technical programs as well students in college preparatory programs.

Although the proposed examination system has superficial similarities to the examination systems in some other countries, there are also fundamental differences. In most other countries there is little concern for issues of statistical comparability of the examinations from year to year or from one region of the country to another. In many cases the examinations of other countries, like the AP examinations, are aimed at only an elite fraction of the students or used for sorting students into educational and occupational tracks. Moreover, few countries have examinations that are truly tied to a national standard, rather they involve loosely-coupled systems of regional examinations controlled by separate boards. Diversity and choice of levels of examinations to match different educational programs have also emerged as examination systems in other countries have expanded to serve a wider range of students.

Of course, it is not just the examination system that is ambitious. The national goals are also quite ambitious. Achieving the goals and the ambitions of the proposed system will require a sustained national commitment. There are a host of political, social, and technical obstacles that will need to be overcome if the vision is to become a reality. Our focus is limited to the technical issues that will need to be resolved in developing the system.

The following sections of this paper identify several major technical issues that need to be dealt with in the development of the system.

VALIDITY

Judgments about the validity of an assessment are concerned with the degree to which the adequacy and appropriateness of the uses and interpretations of assessment results are supported by empirical evidence and logical analysis. In addition to obtaining evidence to support particular interpretations, validation requires an evaluation of the consequences of test

use. Validity is the overarching technical concern for any assessment. Indeed, the other technical issues that are considered in the subsequent sections are important because they contribute to the overall judgment regarding the degree of validity of the assessment.

Intended and Unintended Uses

It is critical that validation plans be built into the design of the nationwide assessment system. Since validity depends on the uses and interpretations of assessment results, a first step in developing plans for validation is the identification and elaboration of the intended uses and interpretations of individual student assessment results and of aggregations of student results to report on schools, school districts, states, clusters of states, or the nation as a whole. To the extent possible, it is also important to identify likely unintended uses and interpretations. Obviously, not all unintended uses and interpretations can be anticipated, but past experience provides the basis for identifying some of the more serious misuses (e.g., tracking into dead-end remedial programs) and misinterpretations (e.g., concluding that failing students or groups of students lack the capacity to achieve the standard based solely on the assessment results).

The Goal 3 Resource Group has already identified some of the intended uses and others are implicit. Among other things, the assessments are intended to help improve American education. Expectations for performance would be made explicit through the assessments. Teachers would be expected to help students prepare for the assessments and students would be encouraged to study for them. In addition, the assessments are expected to provide parents, teachers, and students with information about whether a student's performance meets a national standard. They are also to be used as the means of monitoring the progress of states and the nation toward the achievement of Goal 3. This short list provides a starting point, but a number of more specific questions need to be answered in the development of a validation plan. Some of the types of questions that need answers follow:

1. **Nature of Student Reports.** What will be the nature of the reports to students, parents, and teachers? The national standard language suggests a pass-fail, but finer distinctions might be desired (e.g., fail, pass, pass with honors; A, B, C, D, F). Even if reporting whether a student meets the standard or fails to meet the standard is the primary outcome, students, teachers, and parents may all want to know how far below or above the standard a student's performance falls.

2. **Assessment-Based Decisions About Individuals.** What specific decisions about individual students based on the assessment results will be encouraged and what decisions will be discouraged? For the validation plan, it will be important to know if grade 4 and grade 8 results are intended to be used for purposes such as decisions about providing additional instructional support, tracking, or retention in grade. If used for such decisions, evaluations of the effectiveness of the educational programs to which students are assigned would become an essential part of the validation research.

At the high school level, the stance of the Goals Panel needs to be articulated regarding potential assessment uses such as the award of diplomas, the certification of passing or honors in each of the subject matter areas, admissions decisions by colleges, and hiring decisions by employers. Clarification is also needed with regard to the question of whether a single examination or examinations tailored to the educational programs of students will be used at the high school level. Decisions about such uses will affect specifics of the program such as the provisions that are made for students to re-take the examinations. For example, if the examination is to have important consequences for students when they leave high school and provisions for several possible re-takes are desired, it may be necessary to begin administration of the examination well before the start of the 12th grade. The use of an assessment as the basis for the award of high school diplomas would also demand evidence that students were provided with an opportunity to learn the material required by the assessment. The use of assessment results by employers to make hiring decisions, on the other hand, would require evidence that the assessments either reflect skills required on the job or that the assessment results are related to on-the-job performance.

In order to plan the development of the assessment as well as the validation research, it will be important to know what provisions will be made for assessing children with limited English proficiency. Plans will also require attention to provisions for assessing children with certain handicapping conditions.

3. **Aggregate Results.** Several questions arise in the use of assessment results to report on the performance of schools, school districts, states, or the nation that influence design of research to investigate the validity of the uses and interpretations of aggregate results. For example: What information will be desired at different levels of aggregation (e.g., the percent of students who meet a standard, distributions of scores on assessments)? What decisions will be based on aggregate results for a school or school district? Will aggregated results be separately reported for specific groups of students (e.g., gender or racial/ethnic groups) and, if so, for which groups and for what purposes? How does the use of results at higher levels of aggregation impact on the uses and validity of results for smaller aggregates? Who is to be held accountable for what performance? It would hardly be fair to hold forth grade teachers accountable for the first four years of schooling. Nor would it be reasonable to hold a particular teacher accountable for the performance of students who transfer in and out of several different schools during the course of a school year.

4. **Co-statistics.** What co-statistics (e.g., class size, opportunity to learn, or poverty statistics) about the conditions of education and instructional programs will be desired for reporting? The co-statistics that will be needed will, in turn, depend on decisions about uses of the assessment results.

The answers to questions such as the above will undoubtedly evolve over time. However, it would be useful to begin the discussions necessary for the Goals Panel to be in a position to make the policy judgments necessary to answer such questions. A working group, either the existing Goal 3 Resource Group or a similar group with overlapping membership, could work over the course of the next several months to develop recommendations and simultaneously spell out the types of validation work that would need to be undertaken. Input from representatives from local districts and states regarding potential uses and curriculum related issues would be a vital part of this planning process.

In addition to planning the scope of the validation work, the working group should consider mechanisms for overseeing and conducting the validation studies. Some independent body may need to be established to assure that the effort is credible and consistent with high professional standards.

Needed Validity Studies

Although, as already stated, the details of the validation plan must await more specific answers to questions about uses and interpretations, some general steps can be identified now. A few recommended categories of validation work that will be useful regardless of the answers to the questions are briefly outlined below. A study group could develop preliminary plans for each of these categories during the next several months.

As envisioned by the Goal 3 Resource Group, assessment systems would be developed by clusters of states that come together to define a shared curriculum framework and procedures for assessment. The cluster results would be linked together and to a common national standard, possibly with the use of a national anchor examination. Since the details of use are likely to vary from one cluster to another, more detailed validation plans tied to specific cluster decisions would subsequently need to be developed in conjunction with the partners for each cluster of states.

1. **Judgmental analyses.** Panels of subject matter experts will need to be assembled for purposes of making several types of judgments about the validity of the assessments to be used by any cluster of states. Some examples of the types of judgments that will need to be systematically gathered are (a) the adequacy of the shared curriculum framework for the cluster examination, (b) the alignment of the examination to the shared curriculum framework, and (c) the degree to which the shared curriculum framework is consistent with the national educational standards framework. The curriculum frameworks will also need to be compared to international standards, standards required for next levels of education or employment, and standards of what kinds of knowledge is most useful for transfer and further learning.

   Although existing curriculum frameworks in some states and, in the case of mathematics, the Curriculum and Evaluation Standards for School Mathematics

published by the National Council of Teachers of Mathematics provide excellent starting points, it should be recognized that they are quite general. They lack the level of specificity that is found in detailed course syllabi and, hence, do not provide blueprints for assessments. Nor do they provide an explicit basis for setting standards of performance, that is, the level of performance required to pass or receive honors. Considerable effort will be required to build upon these frameworks to develop a shared curriculum framework for a cluster that is sufficiently specific to create a valid curriculum-driven assessment system that we want teachers to teach to and students to study for.

2. **Analytical studies.** Analyses of scores assigned by raters to student responses to assessment tasks will need to be undertaken. Several types of analyses are likely to needed (e.g., rater calibration, rater drift, inter-rater agreement, generalizability of performance across tasks, calibration of cluster results to a national standard, detection of differential functioning of tasks for subgroups of the population). These analyses are described in subsequent sections of this report, but the results of the separate analyses will all contribute to an integrated judgment regarding the validity of the assessment.

3. **Impact analyses.** A major purpose of the proposed assessment system is the improvement of American education. It is expected that the assessments will clarify expectations and motivate greater effort on the part of students, parents, teachers, and others involved in the educational enterprise. The validation questions regarding the impact of the assessment system ask, not only the degree to which the assessments have these intended consequences, but the nature and magnitude of possible unintended consequences (e.g., student dropout rates, student labeling and tracking).

The impact of the assessments on schools, the curriculum, and the amount of instructional time devoted to different content areas should be evaluated. Since, for example, the type of performance-based assessments that are envisioned by the Goal 3 Resource Group sometimes involve the administration of the tasks to one student at a time, it will be important to evaluate how other students are spending their time while one student is being given an individual performance test by a teacher. Possible changes in the amount of time spent teaching areas not included in the assessment and in the performance of students in those areas should also be documented as part of the validation research effort.

Although important in each of the strands of validation research, the specific uses of an assessment are particularly critical to investigations of the consequential aspects of validity. For example, if it were decided that at grade 4, assessment results would be used to inform teachers, students, and parents of the progress of each student, but not for purposes of assignment to special instructional programs, tracking, or retention, then the study of impact could be relatively modest. The ways in which teachers use the results to plan instruction for individual students and the interpretations of

performance they provide to parents and students could be investigated. Of particular interest would be the identification of uses that benefit students as well as uses that interfere with student progress. Interpretations of results by students and teachers would also deserve study.

The extent and nature of the impact studies would be altered substantially by the introduction of uses with higher stakes for students. For example, if assessment results were used to assign students to a remedial education program, it would be necessary to demonstrate that students assigned to that program were more likely to learn and show greater gains in achievement in that program than in a regular program. Similarly, before the use of an assessment to retain students in grade could be considered valid and defensible, it would be necessary to experimentally demonstrate that students failing the assessment were more likely subsequently to achieve higher standards of performance by being retained than by being promoted.

4. **Sensitivity of results to instruction.** Since the assessments for each cluster are intended to be based on shared curriculum frameworks and provide public statements of what students should study for, the performance of students on the assessments should be sensitive to the degree to which instruction is aligned with the examinations. Comparisons of performance of students with varying degrees of instructional exposure to the shared curriculum would provide evidence that the assessment is sensitive to instructional alignment to the curriculum.

5. **Relationship of results to other student measures.** Evidence regarding both the consistency of the examination results with other indicators of student achievement and the lack of dependency on irrelevant factors will be needed. For example, the relationship of examination performance to ratings of student achievement by teachers and to other existing tests and indicators of student achievement should be investigated. Relationships with measures of characteristics irrelevant to the purposes of the examination (e.g., anxiety, reading ability in English in the case of a mathematics examination) should also be investigated. In addition, the standardized, on-demand, portions of the assessment required by a cluster or a national anchor should be related to more extended interviews and performance measures obtained from long-term student projects and portfolios of work. The latter set of relationships are vital as checks against corruption of the system or the tendency for high-stakes assessments to yield inflated scores of the type that have come to be known as the "Lake Wobegon" effect with current high-stakes testing programs.

6. **Statistical validity of aggregate results.** Representativeness is a critical requirement for drawing valid conclusions about aggregates at the school, district, state, or national level. Census testing, that is the administration of an examination to all students would appear to avoid the question of representativeness. However, even in "census testing" not all students are administered an examination. Students may be excluded from census testing for a variety of quite legitimate as well as less legitimate

reasons. For example, students with limited English proficiency, or students with individual education plans under P.L. 94-142, may be excluded because the examination is judged to be inappropriate for them. Students attending private schools may also be excluded from a state or district mandated examination. Thus, issues of representativeness and comparability from district to district or from state to state within a cluster require attention and analysis even in the case of census testing.

A larger issue of representativeness arises at the national level with the cluster examination model, or, for that matter with a national test, due to the voluntary nature of the system. Census testing of all eligible students in, say, 35 states that voluntarily participate, will not yield nationally representative results. A voluntary national examination system will not satisfy the need of the Goals Panel to monitor progress of the nation unless there is nearly universal participation.

FAIRNESS

As in the case of issues of generalizability and calibration that are discussed below, the issue of fairness of an examination to all segments of the population is fundamental to the overall evaluation of the validity of the assessment system. Fairness is separated from the validity discussion to highlight the importance of the topic.

In America 2000: An Education Strategy, the question of bias against minorities is addressed as follows. "As for bias, the new tests will be screened to eliminate it. Bear in mind that minority parents also want to know how well their children--and the schools their children attend--are doing in relation to the national education goals and standards" (p. 32). Both the issue of screening for bias and the desire for a common standard are relevant to the design of the proposed nationwide assessment system. The issues are considerably more complicated than the brief statement in America 2000 suggests, however.

Traditional Approaches to Avoid Bias

Screening of tasks by judges to eliminate materials that reinforce stereotypes or that are considered potentially offensive to particular minority group members is common practice and would be expected to be applied in the development of both cluster and national anchor examinations. Such screening, while useful, is far from sufficient. It is not just what is screened out, but what is sought for inclusion that needs to be considered in the selection of tasks to assure that the context is relevant to the experience of children from diverse backgrounds.

Questions of potential bias and fairness arise, not only in the selection of tasks, but in the scoring of responses. Training of raters and monitoring of their ratings are critical to assure

that the results are not influenced by factors that are irrelevant to the purposes of the examination. Depending on the examination design (e.g., number of tasks and type of scores), statistical procedures that identify tasks that function differently for minority and majority students, for male and female students, or for students with certain handicapping conditions may also be useful.

Opportunities and Context-Specific Tasks

Traditional task screening activities are necessary components in the design of an assessment system that is fair, but they are not sufficient. This is evident from the basic philosophy of the proposed assessment system. As envisioned, the system is intended to be a means of encouraging all students to prepare for examinations with public standards and expectations. This implies that access to appropriate instructional opportunities and the provision of instructional supports that help prepare students for the assessments needs to be evaluated before the assessment system can be judged to be fair. If, as unfortunately is true now, on average, minority students have less access to the type of thinking curriculum and instructional supports consistent with forward-looking curriculum and examinations expected to be created by each cluster, then the assessment system will not be fair. This suggests the need for systematic monitoring of instructional experiences as well as student outcomes. Because learning reflected on the assessments is cumulative, such monitoring would need to include all grades, rather than only the grades designated the assessments (4, 8 & 12).

Issues of fair access may need to start with analyses of teacher preparation to determine if teachers have had an opportunity to learn about the standards and their implications for instruction. A more complete analysis would require attention, not only to issues of access within the normal school day, but access to private external coaching or test preparation schools that a high-stakes national examination system is apt to foster. Such after-school private schools are quite common in some other countries (e.g., Japan and Taiwan) with high-stakes examination systems.

Although none of the cluster examinations have yet been designed, it has been suggested that some of them may include portfolios of student work as well as projects and on-demand performance examination tasks. Consistent with, or possibly implied by, this array of examination components is the notion that there would be some latitude for choice of tasks or projects by students, possibly in consultation with their teachers.

Choice of tasks also has been suggested as a key for recognizing and respecting cultural diversity and may be a key to achieving fairness. It is argued that providing students with some ability to choose content, context, and tasks within specified limits will decrease the likelihood of disadvantaging students from diverse cultural and social backgrounds. The goal for a cluster examination would be the creation of a variety of functionally equivalent tasks that are specific to a diverse range of cultural and instructional contexts from which a student, possibly with the advice from a teacher, could choose.

The idea of functionally equivalent tasks which would allow students to demonstrate equivalent depths of understanding and problem solving skills on tasks selected to be most compatible with their backgrounds is appealing. But little is known about the feasibility of constructing such tasks. A program of research and development is needed to explore the feasibility, risks, and benefits of approaching the concerns about diversity and fairness in this way.

GENERALIZABILITY

It will be important to provide evidence that the results of the assessment are dependable across raters, the time of rating, and the choice of specific examination questions or tasks. Systematic data collection and analysis procedures will need to be built into cluster rating sessions to evaluate the degree to which the results are dependable. This can be accomplished using existing "generalizability" study designs. At least three specific aspects will need to be built into the generalizability study designs for ratings for any cluster.

Rater Training and Drift

Rater training is critical to the achievement of acceptable levels of inter-rater agreement. As part of the training process, novice raters normally rate benchmark examinations and have their ratings compared to expert ratings that define the benchmarks. Systematic collection and analysis of the ratings of benchmark examinations will need to be designed into the training program and used to establish a criterion of acceptable level of agreement with the expert judgments. After training, periodic checks also need to be built into the rating procedure (e.g., by occasionally including unidentified papers previously scored by experts among the examinations being graded). The previously scored papers provide a systematic check on rater drift and the data needed to take corrective action during the course of the rating work.

Inter-Rater Agreement

The rating design needs to include multiple ratings of a sample of examinations to establish that results are consistent across raters and possibly to determine if each student product requires multiple raters. The specific design for the investigation of inter-rater agreement will depend on the details of the cluster system, but should provide estimates of the level of agreement across raters within levels of the system (e.g., school, district, state, or cluster level) as well as consistency from one level to the next.

Between Task Generalizability

Previous research with performance-based examinations suggests that there is considerable task-specificity in performance. That is, two tasks that are intended to measure similar

understanding or problem solving skills often lead to substantial differences in performance. For example, a student who demonstrates an understanding of electrical circuits by solving one circuit problem may fail to solve a second apparently similar problem while a second student has the opposite pattern of performance on the two tasks. Similarly, student performance on essays written on two different historical topics may depend, not only on the particular topic, but on the interaction of the topic with student background and prior knowledge.

The goal of an examination involving electrical circuits, however, is to generalize, at a minimum, to understanding of electrical circuits and possibly as broadly as the understanding of scientific processes and principles. Similarly, in the case of essays on historical topics, the interest is in generalizing to student ability to analyze historical events and construct reasoned arguments rather than the specific response to a particular task. Hence, it is crucial that the level of generalizability across tasks be evaluated and that this information be used in designing the overall examination.

One approach to dealing with the limited generalizability across tasks is to increase the number of tasks on the examination. This is the common approach with existing multiple–choice tests, but faces feasibility problems when students need an extended period of time to respond to each task. Another approach is to build access to a common prior knowledge base into the examination process itself. For example, students may be given a specific reading assignment prior to the examination and provided with access to resource materials during the examination. Whatever the approach, it is clear that the issue of poor generalizability across tasks will need to dealt with in the design of the examination system.

For each of these generalizability issues (drift, inter–rater agreement, and between task generalizability), systematic data should be collected during the operational conduct of the assessment program. For the latter two issues, it is also desirable to collect and analyze rater and task generalizability data in piloting the system so that the information can be used to design the operational assessment (e.g., determine if all student products should be double rated and how many task samples are needed).

The level of generalizability that will be required will depend on the specific uses that will be made of the results. In general, the level of generalizability required increases as the stakes attached to the results for individual students are increased. For example, a much higher level of generalizability would be required for assessments used for the award of high school diplomas than for assessments used only to inform teachers, students, and parents about a student's current level of achievement. Assuming that the stakes may increase with grade level, it is likely that the generalizability requirements will be least demanding at grade 4 and most demanding at grade 12.

CALIBRATION

The cluster model outlined by the Goal 3 Resource Group allows for each cluster to develop its own curriculum framework and assessment. Although the assessments may vary from cluster to cluster, it is hoped that the results from any given cluster can still be compared to a common national standard through a process of calibration. A set of national anchor examinations constructed to reflect the national standards framework is envisioned and these anchor examinations are expected to serve as the calibration standard against which the various cluster assessments would be evaluated.

There are several possible interpretations of anchor examinations and the way in which they will be used within the cluster model. Clarification of the nature and role of the anchor examinations is needed in order to specify and evaluate possible calibration study designs for the system. It is also worth considering the possibility that the comparability of results from different cluster examinations could be judged against a common national standard without a separate anchor examination.

Two possible types and uses of anchors are described below. Following the description of these two notions of anchor examinations and their role in the system, consideration is given to the possibility that the separate cluster examinations could be compared to a common national standard without the use of a national anchor examination.

Comparisons to Samples from a National Anchor Examination

One possible calibration model is depicted schematically in Figure 1. The model 1 approach would involve the construction of separate national anchor examinations for each grade and subject that would be used for calibration but not for student reporting. According to this model, anchor examinations would be administered, possibly using a matrix sampling design, to a sample of students within each cluster that wished to have its cluster examination compared to the national standard. The cluster examinations would be used for reporting individual student results. The calibration sample would be used only to evaluate the degree to which passing, honors, or other levels of performance established for a cluster examination corresponded to the standards set for the national anchor examination.

By using matrix sampling procedures, it would be possible to administer a relatively large number of national anchor tasks without placing a large burden on any individual student. For example, the students within a participating school might be divided into, say, twenty groups and one or two distinct tasks from the national anchor examination could be administered to each group as part of the cluster examination. For the school as a whole, twenty to forty different anchor examination tasks could be administered in this way. Statistical comparisons of pass rates for the school as a whole would thus be based on a reasonable sampling from the set of anchor examination tasks without requiring any student to respond to more than one or two of the anchor examination tasks.

The application of the first model would not attempt to equate anchor and cluster results for individual students. It would depend on a combination of audits and statistical checks based at the school, district, state, and cluster level comparing the passing rates implied by the matrix sampled anchor tasks with those of the cluster examination. Comparisons of passing rates would provide the basis for evaluating the relative stringency of the anchor and cluster assessments.

Use of Anchor as a Component of Cluster Examination

A second possible calibration model would include the national anchor examination as one component of each cluster examination (see Figure 2). Student results would be based on a combination of the national anchor and the unique examination components that make up the cluster examination model. Only composite results based on the combination of anchor and unique cluster components of the examination would be reported.

A variation of the second calibration model was considered in which the national anchor would not only be a component of each cluster examination, but would be the core that would determine a student's performance relative to the national standard. Results for the unique components of the cluster examination would be reported separately. Although this variation on the second model poses the least difficult technical problems, it is also the least faithful to the concept of separate cluster examinations aligned with the shared cluster curriculum but reported in terms of a common national standard. The unique cluster components might easily become the fringe with most attention focused on the core provided by the national anchor examination. In this case, the model would, in effect, be no different than a single national examination determined, not by the shared cluster frameworks, but by a single national framework--a de facto national curriculum. Hence, this variation on the second model is not recommended.

Linked Cluster Examinations Without an Anchor Examination

A final possible approach that is worthy of consideration would eliminate the national anchor examination entirely. In terms of the schematic in Figure 1, the comparison of the cluster A examination with the cluster B examination would be direct. This might be approached by cross-cluster grading exercises wherein a representative sample of assessment results would be graded by expert teams from other clusters. Representative samples of papers, projects, or other products produced in a cluster assessment would also be subject to audits by teams identified by a National Standards Board. The results of the cross-cluster grading exercises would be exchanged among clusters and used, together with the audit results, as part of a social moderation process to move each cluster toward a common national standard.

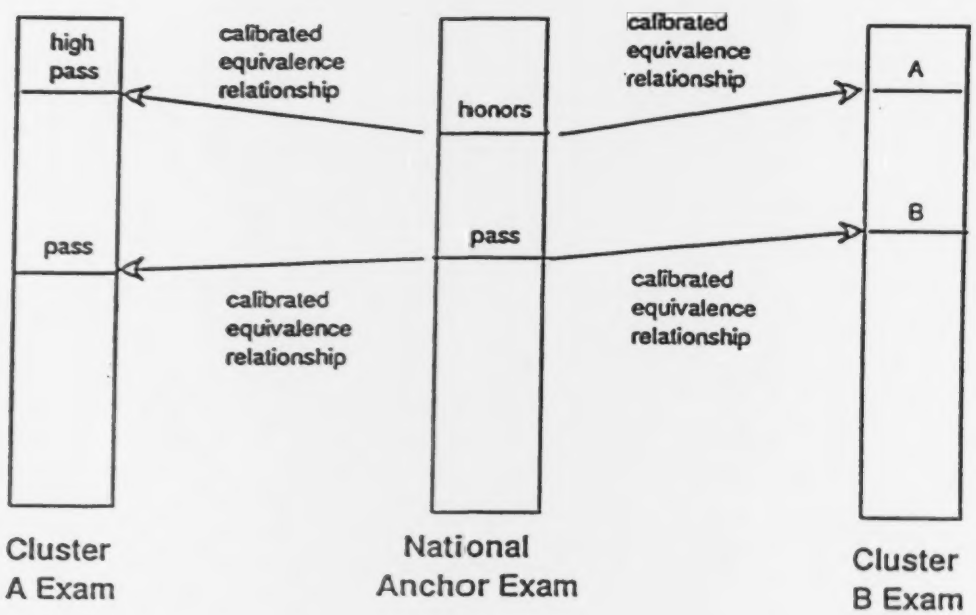
The notion of separate cluster examinations with cross-cluster grading exercises without a national anchor examination is similar to the examination systems found in some other countries. Such a system would not be likely to have as much statistical evidence supporting the comparability of examinations from one cluster to another or from any cluster to a single national standard as one of the two calibration models using a national anchor examination.

On the other hand, such a system is apt to yield examinations that are better reflections of the shared common curriculum of each individual cluster.

Reliance on separate cluster examinations rather than a national anchor examination would have the potential advantage of allowing for multiple approaches to solving the many technical problems that stand in the way of fulfilling the vision of a nationwide assessment system that moves the country toward the achievement of Goal 3. High quality cluster assessments that are publicly visible and shown to have a positive impact on student learning could provide powerful models for teacher training and for commercial publishers of tests, textbooks, and other curriculum materials.

Figure 1

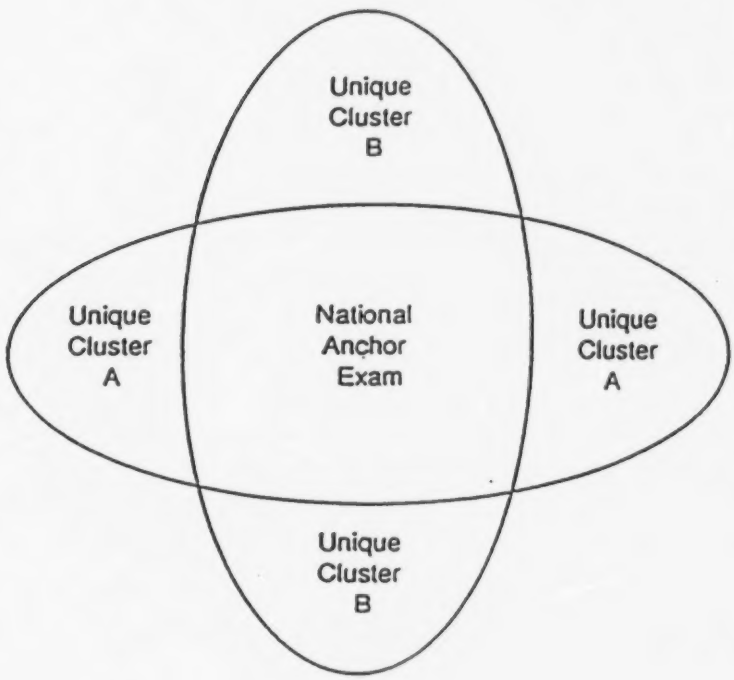
Illustration of Calibration Model 1



G6-17

Figure 2

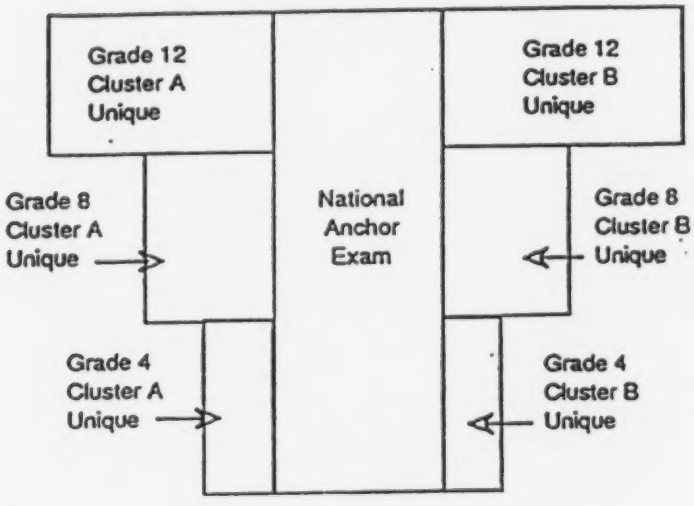
Illustration of Calibration Model 2



G6-18

Figure 3

Illustration of Expanding Cluster Unique Components of Examination with Grade Level



G6-19

CHAPTER 4

GOAL 4: SCIENCE AND MATHEMATICS

GOAL 4: **By the year 2000, U.S. students will be first in the world in science and mathematics achievement.**

Objectives:

- Math and science education will be strengthened throughout the system, especially in the early grades.
- The number of teachers with substantive background in mathematics and science will increase by 50 percent.
- The number of U.S. undergraduates and graduate students, especially women and minorities, who complete degrees in mathematics, science, and engineering, will increase significantly.

GOAL 4 RESOURCE GROUP ON SCIENCE AND MATHEMATICS

RESOURCE GROUP MEMBERSHIP

Convener

Alvin Trivelpiece	Oak Ridge National Laboratory, Tennessee

Members

Iris Carl	Houston Independent School District, Texas
Linda Darling-Hammond	Columbia University, Teacher's College, New York
Edward Haertel	Stanford University, School of Education, California
Kenneth Lay	IBM Corporation, New York
Steve Leinwand	Connecticut State Department of Education
Michael Nettles	University of Tennessee
Senta Raizen	National Center for Improving Science Education, District of Columbia
Ramsay Selden	Council of Chief State School Officers, District of Columbia

GOAL 4 TECHNICAL PLANNING SUBGROUP ON SCIENCE AND MATHEMATICS

TECHNICAL PLANNING SUBGROUP MEMBERSHIP

Leader	
Edward Haertel	Stanford University, School of Education, California
Members	
Leigh Burstein	University of California at Los Angeles, School of Education
Magdalene Lampert	Michigan State University, College of Education
Michael Nettles	University of Tennessee
Penelope Peterson	Michigan State University, Center for the Learning and Teaching of Elementary Subjects
Senta Raizen	National Center for Improving Science Education, District of Columbia
William Schmidt	Michigan State University, College of Education
Ramsay Selden	Council of Chief State School Officers, District of Columbia

September 4, 1991

TO: THE NATIONAL EDUCATION GOALS PANEL

FROM: ALVIN TRIVELPIENCE, CONVENER
THE RESOURCE GROUP ON SCIENCE AND MATHEMATICS

GOAL 4 RESOURCE GROUP

STATEMENT ON THE TECHNICAL PLANNING SUBGROUP'S REPORT

Summary Position of the Goal Four Resource Group on the Goal Four Technical Planning Subgroup Report

Goal 4 states that , **"By the year 2000, U.S. students will be first in the world in science and mathematics achievement."** The three objectives for this goal are:

1. Math and science education will be strengthened throughout the system, especially in the early grades.

2. The number of teachers with a substantive background in mathematics and science will increase by 50 percent.

3. The number of U.S. undergraduates and graduate students, especially women and minorities, who complete degrees in mathematics, science, and engineering, will increase significantly.

Our Interim Report (March 25, 1991) contains five groups of outcome measures related to Objective 1. It also recommends two sets of indicators related to Objective 2. The Technical Planning Subgroup comprised of a group of education experts from Michigan State's Institute for Research on Teaching and the Center for the Study of Evaluation at the University of California at Los Angeles, a National Education Goals Panel staff member, plus four members of the Resource Group.

The Technical Planning Subgroup proposes for additional indicators to supplement the original recommendations of the Resource Group. The Subgroup also specifies additional details regarding some indicators recommended in the Interim Report. Seven topics are included in this group.

All eleven are supported by a methodology statement and, in some cases, sample questions that could be used vis a vis the indicators.

The first set of four new indicators are related to two areas of interest (sse pp. 60–61 of Interim Report).

In addition to these recommendations, the Technical Planning Subgroup specifies further details regarding indicators contained in the Interim Report. These are related to the sections in the Interim Report on *Standards for Curriculum and Learning Goals* (pp.60–61), and *Instructional Conditions* (pp.61–63).

Standards for curriculum and learning goals:

Awareness and adoption of learning goals and curriculum standards by school and educators;

Use of standards and learning goals.

Instructional conditions:

Press for high achievement in mathematics and science for all students;

Organization of instructional time and teachers' time;

Structure of decision making about science and mathematics instruction;

Availability of: high–quality textbooks and associated materials; equipment, expendable, supplies, and laboratory facilities; computers and telecommunications technology;

What is taught and how.

Position on the Technical Planning Subgroup Report

Most of the Resource Group members accept and endorse the Technical Planning Subgroup report. The Subgroup report both supplements and enhances the Interim Report. In one sense their report is a refinement of parts of the Interim Report, especially the section related to objective one.

The Subgroup report also reminds us that, as we develop detailed strategies and indicators to measure progress towards attaining our national goals, we also must be ready to refine our thinking and use all the resources available to ensure success. This will require close interaction between NEGP and many educational organizations both public and private.

One member of the Resource Group expressed mild frustration with the indicators developed by the Subgroup. This individual's preference would be to gather information on a few indicators to help monitor the progress of the reform movement.

As regards the first indicator, on public awareness and acceptance of curriculum standards, it might be more germane to seek opinions on a few selected aspects of reform. These might include questions on the use of calculators all the time and the belief that ALL students can and should study mathematics. Somehow we must understand public perceptions from the view point of parent, student and teacher beliefs, attitudes and practices.

It is important to consider the broad range of things that could be done as we monitor progress towards achieving Goal 4. Yet we must constrain the number of activities and attendant questions for each indicator to those most important concepts and principles.

The Technical Planning Subgroup is aware of the importance of the questions to be asked for each indicator. In fact, their report refers to these questions as "sample questions." Refinements or additions may be necessary. For some proposed indicators no sample questions were suggested and consequently some need to be crafted. The Technical Planning Subgroup proposes several follow-on activities.

In summary, the Resource Group accepts and endorses the recommendations of the Technical Planning Subgroup with the caveats and cautions stated above.

THE GOAL 4 TECHNICAL PLANNING SUBGROUP REPORT ON SCIENCE AND MATHEMATICS

Submitted to

The National Education Goals Panel

through

THE GOAL 4 RESOURCE GROUP ON SCIENCE AND MATHEMATICS

for presentation at

The National Education Goals Panel Meeting

September 4, 1991

GOAL 4 TECHNICAL PLANNING SUBGROUP REPORT ON SCIENCE AND MATHEMATICS

INTRODUCTION

The subgroup met twice, once at Michigan State University, and once in Washington, in order to develop more detailed recommendations regarding long-term data collection and reporting systems needed to monitor national and state progress on Goal 4. The subgroup was chaired by Edward Haertel and included Michael Nettles, Ramsay Seldon, and Senta Raizen; it was augmented by experts from Michigan State's Institute for Research on Teaching (Peterson, Lampert, Schmidt) and from the Center for the Study of Evaluation, UCLA (Burstein) and David Stevenson of the NEGP staff. The purpose of this summary is to propose some modifications of our Interim Report for the whole Goal 4 Resource Group to consider.

In our Interim Report (March 25, 1991), we proposed five sets of outcome measures to measure progress toward the attainment of Objective 1 related to Goal 4, i.e., strengthening mathematics and science education throughout the system. We also recommended two sets of indicators related to Objective 2, i.e., increasing the number of teachers with a substantive background in science and mathematics. We wish to augment these recommendations by adding the following indicators:

- Societal understanding of and support for curriculum standards and learning goals in mathematics and science (see pp. 60-61 of Interim Report):

  Indicator 1:[2] Awareness and acceptance of curriculum standards among the general public

  Method: Addition of appropriate questions to annual Gallup Poll on public attitudes toward education.

  Sample Questions: Are you familiar with the NCTM curriculum standards?

  If yes, what three features do you think best? Worst?

  Has your school/district adopted these standards?

  Indicator 2: Professional development opportunities for teachers

[2] Numbering of indicators is for convenience; the numbers do not indicate priority ordering.

Method: Addition of appropriate questions to SASS and/or NSSME (if this survey will be conducted on a regular basis) in order to elicit information on professional development opportunities that acquaint teachers with the content of standards in math and science and help them implement them in the classroom.

- Direct indicators of teachers' understanding of mathematics and science subject matter and how to teach it:

In making this recommendation, we are not advocating yet another test for teachers.

Indicator 3: Teacher self-reports on their own level of expertise and comfort

Method: Appropriate questions could be included with the questions on staff development (see Indicator 2 above) in SASS or NSSME. The NAEP and TIMSS teacher questionnaires also are appropriate vehicles for eliciting this information.

Indicator 4: Teacher responses to pedagogical problems

Method: Adaptation of vignettes and examples developed at Michigan State University to assess mathematics and science knowledge in the context of realistic pedagogical problems. These could be embedded in the "opportunity-to-learn" components of the teacher questionnaires used in NAEP and for TIMSS.

As well as making these additional recommendations, we wish to specify further details regarding indicators recommended in the Interim Report.

Learning Goals and Curriculum Standards (pp. 60-61 of Interim Report)

In the Interim Report, we stated that learning goals and standards for curriculum should be developed at the national level. The Standards for School Mathematics produced by the National Council of Teachers of Mathematics and Science for All Americans produced by the American Association for the Advancement of Science are exemplars of national standard setting efforts. Once the standards exist, awareness and adoption can be determined through surveys.

Indicator 5: Awareness and adoption of learning goals and curriculum standards by schools and educators

Method: Build appropriate questions into existing surveys (possibilities given in order of preference):

- National Survey of Science and Mathematics Education (NSSME), to be repeated every three years (unfortunately, it is now completely ad hoc)
- SASS Questionnaires (school and teacher level)
- ECS surveys of reform practices and CCSSO indicator project (state level)
- NAEP teacher and school questionnaires

Sample Questions: Similar to those for general public (Indicator 1)

Actual implementation of the standards can be established with any certainty only through classroom observation.

Indicator 6: Use of standards and learning goals

Method: A combination of survey questions and case studies. Survey questions related to the national standards and learning goals in mathematics and science should be included in the teacher and school questionnaires of

- NSSME
- SASS
- NAEP
- TIMSS (this would provide some useful data for international comparisons)

Sample Questions: What instructional blocks or course sequences (in math, science) are offered and taken?

What is taught in these blocks or sequences?

How is it taught--use of appropriate instructional strategies?

Indicate differences among population groups in enrollment in alternative course sequences.

How are decisions made on who enrolls in what sequence?

[Note: Each of these questions needs further specification to elicit the appropriate details. There is considerable experience within all the listed surveys on what sorts of questions provide good information. NELS:88 also included some relevant questions in its first follow-up survey in 1990.]

In-depth case studies of actual classroom and school practice need to be conducted and designed to link to the survey questions.

At the state level, the state level surveys suggested above (ECS, CCSSO) could be used to collect information on existence of reform curriculum frameworks in the states; awareness, adoption, and use could be tracked in the states with the measures recommended for the national level, provided the survey designs include state-representative samples.

Instructional Conditions (pp. 61-63 of Interim Report)

Three areas were outlined in the Interim Report that need to be monitored: a) school organization and structure, b) materials related to the curriculum, and c) instructional practices in the classroom.

a) School Organization and Structure

Indicator 7 : Press for high achievement in mathematics and science for all students

Method: Essentially the same questions and instruments can be used as in Indicator 6 dealing with use of standards and learning goals. The difference is in the analysis and reporting: The goals/standards indicator should report against the requirements of the standards, whereas this indicator should report on learning opportunities in mathematics and science in general, and how these are distributed across student groups, no matter whether the curriculum follows the standards or not.

Sample Questions: What awards are available to students for excellence in math or science?

What special programs or other steps are taken to encourage participation and high achievement in math and science by students from underrepresented groups?

Indicator 8: Organization of instructional time and teachers' time

Method: Appropriate questions to be included in the school and teacher questionnaires of

- SASS
- NSSME
- TIMSS (this would be particularly useful in order to illustrate a wider variety of school organization and use of teacher time than can be observed among U.S. schools)

Sample Questions: What is the school's scheduling structure for a day, a week, a semester or school year?

What are the schedules of the teachers responsible for mathematics instruction? Science instruction?

Indicator 9: Structure of decision making about science and mathematics instruction

Method: Appropriate questions to be included in same instruments as above

Sample Questions: Who decides on curriculum (choices given, e.g., district superintendent, principal, department chair, individual teacher, teacher groups)?

Who decides on textbooks and teaching materials (choices as above)?

How are students assigned to classes/curriculum sequences?

How are teachers assigned to classes/curriculum sequences?

b) Materials Related to the Curriculum

Indicator 10: Availability of: high-quality textbooks and associated materials; equipment, expendable supplies, and laboratory facilities; computers and telecommunications technology

Method: Addition of appropriate questions to existing surveys, as above

Sample Questions: Relevant questions have been included in existing surveys; several past surveys of computer availability (e.g., Becker's at Johns Hopkins) also contain appropriate questions. Adaptations to questions may need to be made to ascertain congruity between standards and available materials.

c) Instructional Practices in the Classroom

Indicator 11: What is taught and how

Method: Survey questions included in existing questionnaires, augmented by in-depth case studies and intermediate-level information consisting of material like teacher logs and time use budgets, analyses of materials used in the classroom (using methodology being developed for TIMSS), and analysis of student work samples (as for NAEP writing assessment and recommended for 1994 NAEP science assessment).

Sample Questions: Same as used for Indicators 6 and 7.

Special attention to how computers and related technology are used.

Next Steps

1. Agreement by Resource Group for Goal 4 to suggested indicators, based on priorities of National Education Goals Panel.

2. Examine existing survey instruments for potentially useful questions, identify those to be used.

3. Develop additional questions where no suitable questions have been identified.

4. Decide whether additional questions are to be included in existing periodic surveys (SASS, NAEP, TIMSS, ECS, CCSSO) or whether a separate science/math specific survey (adapted NSSME) should be conducted periodically, or both.

5. Design and pilot intermediate-level information base on instructional content and practices, linked to related survey questions, as described under Indicator 11.

6. Design and conduct in-depth case studies linked to survey questions on opportunity to learn, instructional content, and teaching practice.

7. State level information on instructional conditions should parallel the national level information. For survey information, the same questions and instruments can be used, provided they yield state-level as well as national information. (This is currently not the case for NSSME or NAEP, unless NAEP continues to go state by state.) This will require:

   - increasing sample sizes to make information state-representative. This would, however, greatly increase survey costs and response burdens (as it stands, there is much more survey information collected than has ever been analyzed);

   - collecting in-depth information for every state through case studies or the like. This does not appear feasible; the intermediate-level information may have to suffice for reporting state-by-state.

CHAPTER 5

GOAL 5: ADULT LITERACY AND LIFELONG LEARNING

GOAL 5: **By the year 2000, every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship.**

Objectives:

- Every major American business will be involved in strengthening the connection between education and work.
- All workers will have the opportunity to acquire the knowledge and skills, from basic to highly technical, needed to adapt to emerging new technologies, work methods, and markets through public and private educational, vocational, workplaces, or other programs.
- The number of high-quality programs, including those at libraries, that are designed to serve more effectively the needs of the growing number of part-time and mid-career student will increase substantially.
- The proportion of those qualified students (especially minorities) who enter college, who complete at least two years, and who complete their degree programs will increase substantially.
- The proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially.

GOAL 5 RESOURCE GROUP ON ADULT LITERACY AND LIFELONG LEARNING

RESOURCE GROUP MEMBERSHIP

Convener		
	Mark Musick	Southern Regional Education Board, Georgia
Members		
	Paul Barton	Educational Testing Service, New Jersey
	Forest Chisman	Southport Institute for Policy Analysis, District of Columbia
	Peter Ewell	National Center for Higher Education Management Systems, Colorado
	Patsy J. Fulton	Oakland Community College, Michigan
	James R. Morris, Jr.	South Carolina State Board for Technical and Comprehensive Education
	William Spring	Federal Reserve Bank of Boston, Massachusetts
	Thomas Sticht	Applied, Behavioral, and Cognitive Sciences, Inc., California
	Mark Tucker	National Center on Education and the Economy, New York

GOAL 5 TECHNICAL PLANNING SUBGROUPS ON ADULT LITERACY AND LIFELONG LEARNING

TECHNICAL PLANNING SUBGROUP MEMBERSHIP

Leader	
Marc Tucker	National Center on Education and the Economy, New York
Members	
Sue Berryman	Teachers College, Columbia University, Institute on Education and the Economy, New York
Paul Barton	Educational Testing Services, New Jersey
Peter Cappelli	University of Pennsylvania, National Center on the Educational Quality of the Workforce
Russell Edgerton	American Association on Higher Education, District of Columbia
Peter Ewell	National Center on Higher Education Management, Colorado
Joy McLarty	American College Testing, Iowa
Thiery Noyelle	Eisenhower Center for the Conservation of/Human Resources, New York
Patrick Terenzini	Pennsylvania State University, National Research Center on Postsecondary Teaching, Learning and Assessment
Daniel Wagner	University of Pennsylvania, National Center on Adult Literacy

September 4, 1991

TO: THE NATIONAL EDUCATION GOALS PANEL

FROM: MARK MUSICK, CONVENER
THE RESOURCE GROUP ON ADULT LITERACY AND LIFELONG LEARNING

GOAL 5 RESOURCE GROUP

STATEMENT ON THE TECHNICAL PLANNING SUBGROUP'S REPORT

Enclosed are recommendations of the Resource Group on Goal 5 regarding the reports of the Technical Planning Subgroup on (1) assessing general education outcomes of collegiate education and (2) assessment of work force skills. As you requested, we have briefly summarized the two reports and provided recommendations on each. We understand that the Technical Planning Subgroup reports themselves will also be available to interested parties.

All Resource Group members received copies of the Technical Planning Subgroup reports. A majority of the Resource Group participated in the teleconference to discuss the reports and a majority of the Resource Group has endorsed the summary statement of recommendations. The record should show that on the recommendations pertaining to workforce skills, Mr. Forrest Chisman requested to be noted as "not participating in the preparation of this statement."

The issue in the two reports that was most debated was that of the staged research and development process for the assessment of workforce skills. The Resource Group did endorse the idea of the staged development process, but the discussion in the group noted the difficulties, costs, time required, and uncertain outcomes and raised questions about priorities and benefits of some of the long range proposals. Tom Sticht's written reaction to the report formed the backdrop for that discussion. I would be glad to talk with you in more detail about the Resource Group recommendations.

ASSESSING GENERAL EDUCATION OUTCOMES OF COLLEGIATE EDUCATION

A Brief Summary of the Report of a Technical Planning Subgroup and Recommendations of the Resource Group on Goal #5 (Adult Literacy and Lifelong Learning) of the National Education Goals Panel

A technical planning subgroup for the National Education Goals Panel's work on Goal #5 (Adult Literacy and Lifelong Learning) was created by the National Education Goals Panel to further consider assessment of the general education outcomes of collegiate education. Briefly summarized, the technical planning subgroup recommended the development of a system that:

1. Assesses a sample of collegiate graduates or seniors about to graduate,

2. Maintains confidentially of results with respect to individual students and individual institutions,

3. Draws samples and reports in such a way as to provide information by types of educational institutions, by the characteristics of the course of study followed by the students, by the majors of students and by the characteristics of the students in terms of their backgrounds, and

4. Reports achievement and proficiency in the context of the achievement levels of the students when they enrolled in higher education.

This assessment system would not be a student or institutional accountability measure. The assessment system would provide one direct measure of the National Education Goals Panel objective pertaining to the proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems.

The assessment system should further:

-- profile graduates on scales that encompass a range of college level achievement and not be established on the basis of minimum level of achievement,

-- use advanced assessment techniques that go beyond multiple choice questions,

-- provide clear communication of what graduates know and can do,

-- provide information to guide the development of national higher education policy, rather than to compare or rank performance,

-- be accompanied by information showing differences in the backgrounds and achievement levels of students entering different types of institutions.

The technical planning subgroup also endorsed the development of supplemental information to this sample-based assessment system. Supplemental information might include outcome indicators on experiences of college students, such as performance on institution-wide writing examinations and other exercises that are required by institutions to earn degrees. There may also be process measures that are seen as an important part of the supplemental information.

Developing a system for assessment of performance of college-level students requires developing a broad consensus among higher education and the public and others about such a system. While such consensus building is time consuming, experience to date in assessment shows that is necessary if the final product is to be successful.

The recommendations of the Resource Group on Goal #5 regarding the report of the technical planning subgroup on assessing general education outcomes of collegiate education.

1. The Resource Group endorses the creation of a system to provide information about American college students' ability to think critically, communicate effectively, and solve problems. The technical planning subgroup's proposal provides an outline on how this might be accomplished and notes the important factors and difficulties that would be part of an effort to create such a system.

2. Neither the Resource Group in its original recommendation nor the technical planning subgroup endorsed the National Adult Literacy Survey as a systematic way to measure collegiate-level critical thinking, communication, and problem-solving. However, the results from the National Adult Literacy Survey will include performance by college graduates who are assessed in the survey. Indications are that college graduates will not show performance that reaches the maximum levels of the National Adult Literacy Survey. This illustrates that the National Adult Literacy Survey can provide information that goes beyond the assessment of basic reading skills. An examination of the scores of college graduates on the National Adult Literacy Survey will provide information in 1993. Results from the National Adult Literacy Survey may serve as supplemental information to the direct collegiate assessment recommended by the Resource Group and the technical planning subgroup, the results of which could not be available before the later half of this decade.

3. In addition to the development of a direct measure of progress toward the objectives of critical thinking, effective communication, and problem-solving, the Resource Group endorses efforts to identify information that supplements knowledge about the collegiate aspect of Goal #5. This supplemental information might include an indirect outcome measure such as the percentage of students

passing institution-wide writing examinations or a process measure such as faculty contact hours as indicated in administrative records.

ASSESSMENT OF WORK FORCE SKILLS

A Brief Summary of the Report of a Technical Planning Subgroup and Recommendations of the Resource Group on Goal #5 (Adult Literacy and Lifelong Learning) of the National Education Goals Panel

The National Education Goals Panel created a technical planning subgroup for Goal #5 (Adult Literacy and Lifelong Learning) to further consider the original recommendation of the Resource Group that the United States Department of Education strengthen the National Adult Literacy Survey and use it to obtain international comparisons of work force skills. The objective of the original recommendation was to indicate to the American public and its leaders "how the functional literacy skills of [the American] work force compare with those of other developed nations." This information was intended to inform the public about the progress on Goal #5 that "every adult American... possess the knowledge and skills necessary to compete in a global economy...."

The recommendations of the technical planning subgroup can be summarized briefly in three main points:

1. The technical planning subgroup supports the idea of international comparisons of work force skills but does not support the idea of using even a strengthened National Adult Literacy Survey alone as the measure of those skills.

2. The technical planning subgroup supports a staged research and development process to build measures of these generic workplace skills and to benchmark them against the skill levels required for expert performance in broadly different occupational categories.

3. The technical planning subgroup recommends that for each country in the cross-national comparisons the sampling process should be such as to include a distribution of those workplace characteristics that affect the skills required of its workers.

The central part of the technical planning subgroup's recommendation is the proposed staged research and development process. This staged process would include four phases.

Stage 1: Extend the National Adult Literacy Survey to handle writing, speaking, listening and mathematics.

Stage 2: Develop and add measures of higher order cognitive skills.

Stage 3: Develop and add measures of orientations toward work.

Stage 4: Develop and add measures of the skills recommended by the Secretary's Commission Assessing Necessary Skills (SCANS).

The technical planning subgroup identified four types of skills and knowledge that it believes broadly constitute the cross-occupational skills required in the workplace:

1. The foundation skills, knowledge, and orientations (these include the basic skills of reading, writing, mathematics, listening and speaking, higher order cognitive skills such as learning strategies and orientation regarding important attitudes such as taking responsibility).

2. The SCANS generic functional skills which appear in many different occupations and industries.

3. Occupationally-specific knowledge and skills.

4. Company-specific knowledge and skills.

The recommendations of the Resource Group on Goal #5 regarding the report of the technical planning subgroup on work force skills:

1. The Resource Group agrees with the technical planning subgroup that a strengthened National Adult Literacy Survey is the first stage in a development process to produce more information about workplace skills. The Resource Group recognizes that as currently defined literacy alone is not adequate for the broader range of knowledge and skills necessary in most workplaces today and in the future. Literacy is a fundamental requirement however. To the extent that the National Adult Literacy Survey can be extended to handle writing, speaking, listening and mathematics that goes beyond simple arithmetic operations, this is an important first step in the assessment of occupational workplace skills envisioned by the technical planning subgroup.

2. International assessments of the knowledge and skills of elementary and secondary students, for example, those in science and mathematics, provide information relevant to measurement and comparison of work force skills. Simply stated, school-aged students of today are the work force of tomorrow. If American

school- aged students are competitive internationally in the areas of science, mathematics, communication and problem-solving, this is one indicator of the skills they may have as adults. Therefore the Resource Group suggests that information from international assessments of student achievement collected in the 1990s be seen as an additional source of information about work force skills. These types of skills were included by the technical planning subgroup in its description of foundation skills, knowledge, and orientations.

3. The Resource Group endorses a staged research and development process recommended by the technical planning subgroup and notes that this is perhaps a decade-long undertaking. This staged development process might produce information not now available and could break new ground in the assessment and comparison of workplace skills. The technical planning subgroup did not support the idea of using even a strengthened National Adult Literacy Survey alone as the measure for international comparisons of work force skills. The Resource Group does not challenge the assertion that the National Adult Literacy Survey alone should not be used in this way. The Resource Group does suggest that a staged research and development process should consider whether it is possible to make a strengthened National Adult Literacy Survey a part of an international work force skills assessment. The investment to date in the National Adult Literacy Survey suggests that while the National Adult Literacy Survey does not encompass the full range of work force skills, literacy is an important component and the measurement capacity of the National Adult Literacy Survey goes beyond basic literacy skills. The National Adult Literacy Survey might play a limited but worthwhile role in some aspects of international assessments.

THE GOAL 5 TECHNICAL PLANNING SUBGROUP REPORT ON ADULT LITERACY AND LIFELONG LEARNING

Submitted to

The National Education Goals Panel

through

THE GOAL 5 RESOURCE GROUP ON ADULT LITERACY AND LIFELONG LEARNING

for presentation at

The National Education Goals Panel Meeting

September 4, 1991

THE GOAL 5 TECHNICAL PLANNING SUBGROUP REPORT ON ADULT LITERACY AND LIFELONG LEARNING

The subgroup strongly endorses two key recommendations of the Resource Group for Goal 5 and makes specific suggestions for implementing those recommendations.

The first recommendation is that the United States devise a method for comparing the skills of American workers with the skills of workers in the other major countries with which we compete. This recommendation gets to the heart of the country's concern over whether our nation has what it takes to compete in an increasingly demanding world. As you will see, we acknowledge the complex issues in conducting such an assessment. Considerable research will have to be done to put together an assessment that will make truly useful comparisons. This is not like an assessment of school learning, because what is at issue is whether workers have the skills required to do particular kinds of work. Because different countries organize work differently, and for other reasons based in cultural differences, it will not be easy to make valid comparisons. No existing instrument will do the job. But it will certainly not be impossible to develop one that can. We lay out what we regard as some eminently practical ways of approaching the task.

The second recommendation is that the nation find out what our graduating college seniors know and can do. Because of the immense diversity of objectives of American colleges, we do not think that the nation should attempt to develop a system for comparing colleges or individuals on a single measure of achievement. We do, however, think that it is very important to sample seniors graduating from their college experience. We lay out a plan for doing just that.

Comparison of Skills

The Resource Panel for Goal 5 proposed that the U.S. Department of Education strengthen the National Adult Literacy Survey (NALS) and use it to obtain international comparisons of workforce skills[3]. The proposal's stated objective is to show the American people "how the functional literacy skills of their workforce compare with those of other developed nations." In the proposal's subsequent discussion, "functional literacy skills" get implicitly equated with workforce skills and skills needed to compete in the global economy.

---

[3] National Education Goals Panel, Measuring Progress toward the National Education Goals: Potential Indicators and Measurement Strategies. Compendium of Interim Resource Group Reports. Washington, DC: U.S. Government Printing Office, March 25, 1991, p.93.

Summary Recommendations

The Technical Panel supports the idea of international comparisons of workforce skills, but does not support the idea of using even a "strengthened" NALS alone as the measure of those skills. NALS is a seriously limited measure of the generic skills required in workplaces, and its levels are not benchmarked against the levels required for expert performance in broadly different occupational categories.

The Technical Panel supports a staged research and development process to build measures of these generic workplace skills and to benchmark them against the skill levels required for expert performance in broadly different occupational categories.

For each country in the cross-national comparison, the Technical Panel recommends sampling the country's workplaces to describe the distribution of those workplace characteristics that affect the skills required of its workers. Employer skill requirements affect the skills that workers develop and maintain. The distribution of workers' skills cannot be interpreted out of context of employers' skill requirements.

Backup Discussion

The Technical Panel recommends that any cross-national comparison include the cross-occupational skills required in the workplace: the foundation skills and SCANS skills. Worker performances involve four types of skills and knowledge:

- the foundation skills, knowledge, and orientations (these include the basic skills of reading, writing, mathematics, listening, and speaking; the higher order cognitive skills, such as metacognitive skills and learning strategies; and attitudes or orientations, such as taking responsibility);
- the SCANS generic functional skills, which appear in many different occupations and industries;
- occupationally-specific knowledge and skills; and
- company-specific knowledge and skills.

The Panel felt that K-12 should have responsibility for developing the foundation and SCANS skills. These skills therefore properly fall within the purview of the National Goals Panel. The educative responsibility for occupationally-specific skills is shared between K-

12, post-secondary schools, and employers, but the very specificity of these skills make them unlikely candidates for a cross-national assessment[4]. Company-specific skills are proprietary, entirely the training responsibility of employers, and therefore outside the scope of the National Goals Panel.

A cross-national assessment of the foundation and SCANS skills requires a measurement battery that extends substantially beyond even a "strengthened" NALS. NALS measures "adult literacy skills" (reading and numerical operations). These are part of, but are not the same as, generic, cross-occupational "workplace skills." The NALS mathematics measures are restricted to numerical operations; it does not measure writing, speaking, and listening skills; it does not measure the higher order cognitive thinking skills; it does not measure attitudes or orientations important in the workplace; it does not measure the SCANS skills.

To develop an adequate cross-national assessment of cross-occupational workplace skills, the Technical Panel recommends a staged development process.

- Stage 1: extend NALS to handle writing, speaking, listening, and mathematics. (It now handles only arithmetic operations and does not cover the broader domain of mathematics needed in the workplace.)
- Stage 2: develop and add measures of higher order cognitive skills.
- Stage 3: develop and add measures of orientations toward work.
- Stage 4: develop and add measures of the SCANS skills.

The Technical Panel recommends that even the R&D for the cross-national assessment be done cross-nationally, preferably through a cross-national R&D team drawn from research institutes in the different countries. Measures need to incorporate the range of important variations found in countries that we ultimately expect to involve in the full assessment. Cross-national comparability problems have to be confronted. And if the R&D work is done through a cross-national R&D team, the R&D process itself builds credibility and support for the full assessment within the countries in which we hope to conduct it. It is not clear, when, if ever, either the R&D or the assessment itself have to involve government-to-government contacts.

---

[4] They would almost certainly make the assessment impossibly cumbersome. Measuring them cross-nationally also requires roughly comparable occupational structures across countries, a condition that may not be met.

The Technical Panel recommends that the levels of each of the skills measured in the cross-national assessment be benchmarked against the levels required for expert performance in broadly different occupational categories. We need to know what levels of competence in the different foundation and SCANS skills are required for expert performance in broadly different occupations. For example, some empirical work suggests that jobs can require fairly low levels of literacy, but quite high levels of problem-solving. Without this benchmarking, we cannot relate different levels of performance on the different foundation and SCANS skills to different levels of job performance in broadly different occupations.

The Technical Panel recommends obtaining a descriptive distribution by country of those workplace characteristics that affect skill and skill level requirements. This gives us the distribution of skill demand and use within a country, which provides context for interpreting the distribution of workers' skills. The workplace, not just the occupation or job, affects workplace skill requirements. The same occupation in a small versus large firm, or in a firm organized for mass versus flexible production, will be structured differently. In other words, workers' skills (both skill types and levels) should be evaluated relative to the organizational context for their use.

The details of sampling and data collection strategies should be left to the R&D teams. However, the Technical Panel noted that companies or households could be used as the sampling unit to measure both skills and workplace characteristics. Or two separate studies could be conducted. One would sample companies to get the distribution of workplace characteristics that affect the skills needed. The second would sample individuals to measure skills, perhaps with a brief profile of the individual's workplace.

Indicators of General Education Outcomes of College Education

SUMMARY

The principal recommendation is for the development of sample based indicator(s) of the general education outcomes of higher education (as defined in the National Goals). This would be necessary if the Goals Panel wishes to track progress toward the fifth objective of goal #5. Appropriate context information would also be collected. The participant institutions, selected for the sample, would not be identified, nor would the students. Thus this would not be an "accountability" measure. While this is the recommended direct measure, the Subgroup also suggests the development of supplemental information on processes, as

well as a few "proxy" outcome measures that could be in place more quickly, from information now recorded by higher education institutions.

Introduction

The task addressed here is the creation of a means by which the annual "report cards" of the National Education Goals Panel can track progress toward an objective of its goal #5, that the proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially.

The Interim Report of the Resource Group on Goal 5, dated March 25, 1991, pointed out that to track progress toward this objective "a new kind of assessment will have to be created." The Goals Panel created this Technical Planning Subgroup to, among other things, explore the options for such a new assessment, and recommend an approach.

The objective itself has already been established, as has the decision to develop the best possible reporting system to track progress toward all six goals adopted by the President and the Nation's governors. Starting with these decisions, the Subgroup considered the matter of the most feasible approach to enabling the Panel to report to the nation.

General Education Outcomes

The objective of the Panel is obviously predicated on a belief that there are expected general[5] outcomes of a college education, beyond knowledge within specific subject matter, despite the

a. diversity of institutions and major programs

b. diversity of views about the purposes of higher education

c. diversity of students, in terms of backgrounds, age, objectives, and abilities as they enter higher education.

While consensus on the matter is incomplete, a decade of serious attention to the issue of assessment of outcomes in higher education is generally supportive of the reasonableness of expecting higher education to increase students' ability to think critically, communicate effectively, and solve problems. A considerable number of states, individual institutions, and test development organizations have approached the assessment issue from a similar perspective.

---

[5]We recognize that the term "general education" has taken on specific meanings in some settings. Here, we mean no more than the statement of the Goals Panel.

At the same time, the Subgroup underscores the limited nature of this reduction of higher education aims. The outcome stated by the Goals Panel comes nowhere near to capturing the full value of a four-year college education, nor does it resonate the uniqueness claimed by Ivy College or State University. However, this is not an argument against the usefulness of having indicators based on consensus about higher education's most common denominator -- nor of starting a process which may engender consensus about even larger purposes.

The Context of Thinking and Problem Solving

While the common elements of a successful college education may be worthy of measuring for progress, the more removed they are from the contexts in which they were achieved and applied the more they will be disembodied and begin to resemble tests of verbal ability or general "intelligence." These desired abilities are imparted in subject matter contexts, and they are expressed by students in their successful negotiation of the material which make up an academic discipline. Justice would more nearly be done to students and institutions by assessing students' ability to think critically, etc. within the disciplines that have formed their course of study. To do so, of course, is beyond the realm of practicality, at least at the point of getting started.[6]

Nevertheless, it would be a mistake to abstract from the disciplines entirely; an assessment should aim to be somewhere between a set of generic (all purpose) skills and those that are embodied in performance within a subject or a discipline. Arguments will be advanced for both extremes (for example, a proposal was made recently to use the SAT or an intelligence test as an outcome measure for higher education; those committed to a discipline bound view of achievement will argue for discipline specific assessments). The tension has generally resulted in taking a middle course by virtually all new assessment systems. Examples of anchoring generic skills within broad disciplinary areas are the following:

- The College Outcome Measures Program (COMP) and the Collegiate Assessment of Academic Proficiency (CAAP), offered American College Testing Program
- The Academic Profile, offered by Educational Testing Service

[6]Of course, subject matter knowledge in a major can be assessed, and such assessments are now used in higher education institutions. The Office of Research of the Department of Education has proposed to implement 3 of the 5 models of summative baccalaureate assessment previously developed and reported in Signs and Traces (1990), starting in 1992. These will cover 3 disciplines in the sciences and applied sciences.

- The General Intellectual Skills Test of The New Jersey College Outcomes Program

- College–Base at the University of Missouri

An example would be the assessment of reading, writing, and critical thinking in the areas of humanities, social sciences, and national sciences, used by one of these assessments.

We recommend that a similar course be followed for a Goals Panel assessment. The choice of specific skills and contexts should evolve in the development phase of an assessment system and be informed by the experience of the decade in creating outcomes assessment in higher education. No existing assessment is likely to satisfy the broad consensus required for indicators of progress toward a national goal.

The Indicators Approach

The Subgroup has not reached beyond the Goal, or the Resource Group recommendation. The establishment of a national goal and the call for an indicator to measure progress is itself a momentous step. It will require resources, time, and considerable resolve to bring the recommendation to fruition.

We recommend that this focus on outcomes indicators be maintained. By this we mean the creation of a system

- that assesses a sample of graduates or seniors about to graduate, and institutions these students attend

- that maintains confidentiality of results both with respect to students and institutions (However, samples could be so designed, or augmented, so as to

  provide information to a participating institution for its own use, if it so desired)

- that draws samples and reports in such a way as to provide information by

  -- the type of the educational institution, such as large research universities, small liberal arts colleges, four years versus two year degrees, etc.

  -- the characteristics of the course of study followed by the students, and their majors

-- the characteristics of the students in terms of their backgrounds, their race/ethnicity, gender, whether they are U.S. or foreign students, and their skills and abilities when they entered school from such tests as the SAT and ACT

- that achievement and proficiency be reported in the context of the ability levels of the students when they enrolled, so that the option exists for taking into account these differing entering ability levels.

The Subgroup recognizes that there are a number of objectives that are being, and can be, pursued in the assessment of higher education outcomes. This recommendation of an indicators approach is not intended to be a substitute for other legitimate purposes of assessment, and will not serve, for example, as a system for individual or institutional accountability. It is not offered as a substitute for assessments now in use. Other approaches to an assessment system discussed by the Subgroup are summarized below.

1. Develop a standard of general education achievement for all higher education graduates, and assessments that would measure the degree to which the standard was reached (this is in contrast to a proficiency scale on which students are arrayed, as is used in NAEP). This greatly over-reaches the stated objective of the goals panel; it is not recommended by this Subgroup.

2. Use the Adult Literacy Survey of 1992, (modified, or not). This is a household assessment that locates adults on three literacy scales: prose, document, and quantitative. As did the Resource Group, we questioned whether "literacy" should be the basis for an outcomes assessment of higher education, although the 1992 study results will provide useful information for adults at all levels of education, and will provide useful supplemental information.

3. Ask groups of higher education institutions with similar goals to combine to create an assessment themselves. Aim for economies of scale. This contemplates several or many different assessments and leaves the initiative largely within the higher education community. There is no reason to suppose that higher education institutions would act on their own to create the tracking system desired by the Goals Panel.[7]

[7]A consortium approach is used in Tennessee, but involved a very small number of institutions, with few others interested.

Quality Assessment, Based on Consensus

The objectives of higher education are too important to trivialize with a simplistic assessment, quickly designed and fielded to get a number into a national report card. This will injure the outcomes the Panel wants to advance. Neither can an assessment be developed in isolation and imposed by simple edict or legislative act. To be useful, used, and constructive, the more difficult path of developing a broad consensus among higher education, the public, and affected constituencies, will have to be followed.

Considerable experience to date in state-based higher education assessment programs convinces us that consensus-building is time-consuming, but that it is well worth the time invested. Indeed, many states have reported that the goal development process inherent in consensus-building can have an immediate, positive effect on institutional behavior before any data are actually collected. A system of sample based outcomes indicators should have these attributes:

- It should profile graduates on scales that encompass the full range of college level achievement, and not be of a basic or minimum level of achievement. While it would deal with this "advanced" level of achievement, it should not attempt to dictate a single standard of acceptable achievement

- It should use advanced assessment techniques that go beyond customary multiple choice questions, and include constructed responses, performance tasks, essays, and possibly even portfolios of actual work.

- The system should provide clear communication of what graduates know and can do, rather than simply scores disembodied of meaning. This would include descriptions of achievement at various points along the scales, as well as representative items on the assessment for each level.

- The system should provide information to guide the development of national higher education policy; it should not be used to compare or rank performance.

- Comparisons of scores across institutional types should be accompanied by information showing differences in backgrounds and abilities of students entering each type.

Supplemental Information

In addition to development of a direct measure of progress toward the objective, there are less direct indicators that could be encouraged by the Goals Panel, that could be collected from

records now available or added to surveys, and that could serve to encourage educational progress in higher education.

An example of a proxy outcome indicator might be, for example, the percent of students who passed an institution-wide writing examination (of upper level writing) by the time they enter the senior year. Another might be the percent of students who have had a degree-qualifying exercise (thesis, exhibit, exam, portfolio) in which faculty judgment was rendered.

Examples of a process measures might be faculty contact hours (from institutional administrative records), the requirements for a baccalaureate degree, or the amount of coursework in selected areas (from analysis of student transcripts). A paper by Peter Ewell and Dennis Jones is in preparation on the possibilities for such measures.

Further Elaboration and Development

The Subgroup has not attempted to function as a design team. It has explored approaches that would enable the National Education Goals Panel to track progress toward the

goal/objective established by the President and the Governors. It has elaborated the sample-based assessment approach that would most directly provide such an indicator, without addressing the matter of individual or institutional accountability (and this report conveys neither agreement or disagreement with such a national system of accountability).

The Subgroup emphasizes that however important and desirable it is to have the objective, and a measure to track progress toward it, as it is written it touches only one aspect of the many important purposes of higher education. This report is not meant to endorse a narrowness of purpose.

We are aware that opportunity will exist in just a few months for an elaborated discussion of the matter of creating indicators, through the papers and conference now being arranged by the National Center for Education Statistics. The papers for that effort are being commissioned now, and the conference is expected in early Fall. The purpose of the conference is to further the ability of the Panel to measure progress toward this higher education objective. Current plans are for papers to be prepared on the following topics:

1. Measurement and Timing Issues

2. State Experiences

3. Institutional Experiences

4. Relationship to Pre-Collegiate Testing

5. Testing Services Experiences

6. New Jersey and Other State Tests

7. Relationship to NAEP and the Adult Literacy Survey

8. Job Skills Issues

9. Indirect Assessment Approaches

This effort of NCES should provide the Panel with an excellent opportunity to advance the creation of an indicator of higher education outcomes.

CHAPTER 6

GOAL 6: SAFE, DISCIPLINED, AND DRUG-FREE SCHOOLS

GOAL 6: **By the year 2000, every school in America will be free of drugs and violence and will offer a disciplined environment conducive to learning.**

Objectives:

- Every school will implement a firm and fair policy on use, possession, and distribution of drugs and alcohol.
- Parents, businesses, and community organizations will work together to ensure that schools are a safe haven for all children.
- Every school district will develop a comprehensive K–12 drug and alcohol prevention education program. Drug and alcohol curriculum should be taught as an integral part of health education. In addition, community-based teams should be organized to provide students and teachers with needed support.

GOAL 6 RESOURCE GROUP ON SAFE, DISCIPLINED, AND DRUG-FREE SCHOOLS

RESOURCE GROUP MEMBERSHIP

Convener		
	John W. Porter	(formerly) Detroit School District, Michigan
Member		
	C. Leonard Anderson	Portland School District, Oregon
	Constance E. Clayton	Philadelphia School District, Pennsylvania
	Delbert G. Elliot	University of Colorado
	Joseph A. Fernandez	New York City School District, New York
	Michael Guerra	National Catholic Education Association, District of Columbia
	J. David Hawkins	University of Washington, Social Development Research Group
	Lloyd D. Johnston	University of Michigan, Institute for Social Research

GOAL 6 TECHNICAL PLANNING SUBGROUP ON SAFE, DISCIPLINED, AND DRUG-FREE SCHOOLS

TECHNICAL PLANNING SUBGROUP MEMBERSHIP

Leader

Lloyd Johnston	University of Michigan, Institute for Social Research

Members

Janet Collins	Centers for Disease Control, Division of Adolescent and School Health, Georgia
Fred Newman	University of Wisconsin, Center on Organization and Restructuring of Schools
Barbara Schneider	University of Chicago, National Opinion Research, Corporation, Illinois
Gary Wehlage	University of Wisconsin, Center on Organization and Restructuring of Schools

September 4, 1991

TO: THE NATIONAL EDUCATION GOALS PANEL

FROM: JOHN PORTER, CONVENER
THE RESOURCE GROUP ON SAFE, DISCIPLINED, AND DRUG-FREE SCHOOLS

GOAL 6 RESOURCE GROUP

STATEMENT ON THE TECHNICAL PLANNING SUBGROUP'S REPORT

The Goal 6 Technical Planning Subgroup reviewed the long-term indicators proposed by the Resource Group for Goal 6 and tried to provide a set of recommendations regarding specific measures to be used for each. In the process it also suggested a slight reorganization of the indicators for the third element of the goal "disciplined environment conducive to learning." Previously two indicators were proposed: student misconduct and perceptions of disruption impeding learning. It was felt that these were difficult to distinguish conceptually and should be merged into one indicator called "student misconduct which impedes learning." Measures are suggested for two distinguishable elements of this indicator: disruption in the classroom and truancy/tardiness. The later element was originally considered by the Resource Group and mentioned in its report as an option. It was subsequently considered by the Panel, which seemed to favor its inclusion. The logic for its inclusion is that excessive tardiness and absenteeism not only is disruptive of learning for the student exhibiting these behaviors, but also influences other students by disrupting and slowing down progress in the classroom.

While it was proposed that these two indicators be merged, it was also proposed that a question about rule enforcement be removed from them and placed with a couple of other questions. In combination these questions might comprise a new indicator called "enforcement of rules."

Regarding the element of Goal 6 entitled "schools free of violence and crime," the Technical Planning Group suggested the inclusion of a third indicator (in addition to "victimization at school" and "feeling safe at school") which deals with "carrying weapons at school." It would require only one question to be added to the two student data sources.

The two existing survey series recommended by the Resource Group for measuring progress on Goal 6 – the Monitoring the Future study and the Youth Risk Behavior Surveillance System –

have both indicated a willingness to add items and to provide the Panel with the relevant data from existing items chosen by the Resource Group.

In May 1991 the Resource Group listed ten possible indicators for the September 1991 Report. The Technical Planning Group Report identifies nine indicators that might be used for the long-term data collection.

Indicators 8, 9 and 10 from the May 1991 document were dropped and two new indicators are added by the Technical Group. Those are (1) Sale or Distribution of Drugs at School and (2) Enforcement of Rules.

The Technical Planning Subgroup's seven-page report was divided into three elements. Element I contained four recommended long-term indicators. Element II contained three recommended long-term indicators and Element III contained only two recommended long-term indicators. Listed below are the issues:

Issue 1: Following on the general sentiment of the Resource Group, the Technical Planning Subgroup did not recommend the collection of data to measure the success of the three Instrumental Objectives adapted by the Education Goals Panel.

THE RESOURCE GROUP CONCURS

Issue 2: The Technical Planning Subgroup Report recommended only two new indicators (a) Sale or Distribution of Drugs at Schools and (b) Enforcement of Rules.

THE RESOURCE GROUP CONCURS

Issue 3: In several instances, namely in terms of (a) drug use, (b) being under the influence, (c) feeling safe at school and (d) carrying weapons in school, there was no change in the indicator, but the recommendations were more specific as to the types of questions to be asked.

THE RESOURCE GROUP CONCURS

Issue 4: In Element III, the Technical Group is recommending that Indicator I combine previous indicators which distinguished behaviors and perceptions under one indicator of Student Misconduct which impedes learning. There would be two components under this indicator: Disruption in the classroom and truancy/tardiness.

THE RESOURCE GROUP CONCURS (BUT NOTES THAT THESE ARE SEPARATE MEASURES).

In summary, the Technical Planning Subgroup Report highlights five issues:

- First, there are no recommended indicators to measure the three objectives;
- Second, there is a recommendation that to the extent possible, selective measures be taken at the school level to permit analysis which look at the outline measures as a function of type of school;
- Third, that a measure on Sale or Distribution of Drugs at School be added;
- Fourth, that two separate Indicators on Student Misconduct Which Impedes Learning be merged together; and,
- Fifth, that a new Indicator on Enforcement of Rules be added.

The Resource Group is in support of each of these recommendations as noted above.

THE GOAL 6 TECHNICAL PLANNING SUBGROUP REPORT ON SAFE, DISCIPLINED AND DRUG-FREE SCHOOLS

Submitted to

The National Education Goals Panel

through

THE GOAL 6 RESOURCE GROUP ON SAFE, DISCIPLINED AND DRUG-FREE SCHOOLS

for presentation at

The National Education Goals Panel Meeting

September 4, 1991

THE GOAL 6 TECHNICAL PLANNING SUBGROUP REPORT ON SAFE, DISCIPLINED, AND DRUG-FREE SCHOOLS

SUMMARY

Members of the Goal 6 Technical Planning Group met for the first time in Madison, Wisconsin to develop more detailed recommendations regarding the long term data collection and reporting systems needed to monitor national and state progress on Goal 6. Conclusion reached during that meeting are presented here for the purpose of assisting the Resource Group in its future deliberations.

To begin, a general discussion was held about the appropriate level of analysis for Goal 6. That is, should data be presented at the individual level (e.g., the percent of students nationwide who have tried tobacco) and/or at the school level (e.g., the mean percent of students who have used tobacco by school)? It was noted that a mean of school averages can be deceptive in that the very small schools weigh in equally with the very large schools and thus can dominate the average; but the group felt that it is appropriate to examine the degree of variability across schools, as well as to examine differences between different types of schools. It is anticipated that both Monitoring the Future (MtF) and the Youth Risk Behavior Surveillance System (YRBSS) can be analyzed at the school level. Whether SASS can be analyzed at school level has not been determined. However, the degree of accuracy of school-level estimates is yet to be estimated. Confidence intervals could be unfavorably large, particularly for the YRBSS.

The workgroup recommends that a minimal, standard, set of school-level information be gathered regarding participating schools to include: type of school (e.g., ungraded, alternative, technical, elementary, middle, high), region, urbanicity, grade structure, enrollment size, race/ethnicity breakdown, and socio-economic status (using proxy measures such as the percent of students receiving free lunches, the percent of students who go on to college, and parent education).

Lloyd Johnston indicated that MtF has the flexibility to add all of the additional questions suggested here and below, and Janet Collins indicated that YRBSS could add up to a dozen questions total to that study. This means that the state level indicators will have to be a subset of the national indicators used.

Element I: Drug Free Students and Schools

Indicator 1: Student Drug Use

Age/grade of first use, prevalence of use during lifetime, and prevalence of use during the past 30 days (*highest priority) should be reported for the following drug categories:

A. Alcohol
B. Tobacco
C. Marijuana
D. Cocaine
E. Crack

Beyond these drug categories, what can be reported from the two studies differs. YRBSS has a single residual question which reads:

> *During your life, how many times have you used any other type of illegal drugs such as LSD, PCP, ecstasy, mushrooms, speed, ice, heroin or pills without a doctor's prescription?*

MtF can provide national data comparable to that suggested for the five drugs above on an additional 8 drug categories:

F. Stimulants (amphetamines)
G. Sedatives (barbiturates)
H. Tranquilizers
I. Heroin
J. Narcotics other than heroin
K. Hallucinogens [LSD and PCP could be given separately]
L. Inhalants
M. Steroids

To the extent that a more consolidated measure of illicit drug use in lifetime or past 30 days is needed, the following indexes also would be available at the national level from MtF:

N. Use of any illicit drug
O. Use of any illicit drug other than marijuana

The MtF data are available at the 8th, 10th, and 12th grade levels as of 1991 and the YRBSS data at the 9th through 12th grade levels for the nation as a whole. The same data are available for selected states through the YRBSS. These data can be presented in table form with the percentage of students reporting drug use by age, sex, and race/ethnicity. If possible,

some types of school level data should be reported to characterize differences between schools in amount of student use by type/location of school.

Indicator 2: Peer Norms

Perception of the extent to which peers disapprove of drug use has been found to be an important predictor of drug use. MtF can provide information on the degree of perceived disapproval by peers of using various substances at various levels. State level data would not be available on this indicator via the YRBSS due to its exclusive focus on risk behaviors.

Indicator 3: Being Under The Influence of Alcohol and Other Drugs at School

For drugs and alcohol, the critical measurement feature is whether students are attending class while "high," a condition which may seriously impair their ability to learn. Indeed, even if relatively little use of drugs or alcohol occurs on school grounds, drinking or using drugs prior to school, or at lunch, can affect performance hours later. However, for tobacco we believe that measurement of use on school grounds, at school sponsored activities, or on school buses is most appropriate, but perhaps of lower priority.

This indicator is not currently measured by MtF, or YRBSS. We recommend that three items, such as those listed below, be added to both of these measurement systems. Possible wording was proposed; however, we assume that any new items would be refined and field tested before use.

A. During the past week has there been any time when you were at school and under the influence of alcohol (regardless of where you may have consumed it)?

B. During the past week has there been any time when you were at school and under the influence of any illegal drug (regardless of where you took it)?

C. During the past week have you smoked cigarettes or used chewing tobacco while at school?

Additionally, we recommend the inclusion of an item on the Department of Education's Schools and Staffing Survey (SASS) to obtain information from teachers about the numbers of students who are under the influence while attending class. For example:

> During the past week, how many different students do you think were under the influence of alcohol or drugs while attending your classes?

Indicator 4: Sale or Distribution of Drugs at School

The workgroup believes that students will be unlikely to admit dealing or distributing drugs. Furthermore, while "connections" are likely to be made at school, it is considerably less likely that an actual drug transaction will occur on school grounds. Some of the ideas that were proposed to measure this indicator via the MtF and YRBSS are presented below. However, the workgroup feels that this indicator is of low priority due to complex measurement issues and the anticipated low prevalence of dealing on school grounds.

1. How difficult do you think it would be to buy or get an illegal drug at school, if you wanted one?

or

2. How difficult do you think it would be to buy or get an illegal drug from someone who attends your school?

or

3. In the past year has anyone sold or given you an illegal drug while you were at school? [Preferred]

or

4. In the past year has anyone who attends your school ever sold or given you an illegal drug?

Offering to give or sell drugs might also be covered by a question parallel to whichever of the questions above is chosen.

Element II: Schools Free of Violence and Crime

Indicator 1: Victimization at School

MtF now contains seven victimization items relating to events that happened at school or on a school bus. These should all be used at the national level. Four were deemed to be the most important by the workgroup (threat of assault without a weapon, threat of assault with a weapon, actual assault without a weapon, actual assault with a weapon) and it is recommended that these four be included in the YRBSS state surveys. Responses to these items could be reported separately or as a single index. (The other items in MtF deal with vandalism of one's property, theft of something worth under $50, and theft of something worth over $50.)

Indicator 2: Feeling Safe at School

Fear of victimization may contribute to truancy, dropout, and poor school attendance. The workgroup recommends that two items be added to the MtF and the YRBSS to measure feelings of safety. One item would address feeling safe on the way to or from school. (Revised items based on the School Crime Supplement, 1989, may be helpful here.)

1. How often do you feel unsafe (that is, fearful that someone might physically harm you) when you are at school or on school grounds?

2. How often do you feel unsafe (that is, fearful that someone might physically harm you) when you are on your way to or from school?

In addition, it is recommended that the same two items be added to SASS to examine teachers' feelings of safety at school.

A third item to consider for both students and teachers relates to behavioral consequences:

3. In the past month, did you ever stay home from school because you felt you would be unsafe at school, or on your way to or from school?

Indicator 3: Carrying Weapons to School

The workgroup recommends that a single item be added to MtF and the YRBSS such as,

> "During the past month, did you carry a weapon (such as a club, gun, or knife) to school on one or more days?"

Element III: Disciplined Environments Conducive to Learning

Indicator 1: Student Misconduct which Impedes Learning

The workgroup recommends measuring the amount of student misconduct by examining the number of occasions or amount of time that teachers must deal with discipline problems. Thus, an item such as the following is recommended for inclusion on SASS.

> On average, for every five classes you teach, how many times do you have to interrupt your teaching to deal with student disruption or misbehavior?

As companion measures it is recommended that items such as the following be added to MtF and YRBSS:

In an average school day about how many times does your teacher have to interrupt the class to deal with student disruption or misbehavior?

How often does misbehavior in class by other students get in the way of your learning?

In addition, it is recommended that student reports of unexcused absences, cutting classes, or tardiness should be obtained by adding the first two items below (which are from MtF) to YRBSS, and adding the third item below to both studies:

1. During the last four weeks, how many whole days of school have you missed because you skipped or "cut?"

2. During the last four weeks, how often have you gone to school, but skipped a class when you weren't supposed to?

3. How often do you come to class late (after class has begun) without an approved excuse?

As a companion measure it is recommended that items such as the following be added to SASS:

1. In an average class day about how many of your students are absent from class without an approved excuse even though they are at school that day?

2. In an average class day about how many of your students come to class late without an approved excuse?

Indicator 2: Enforcement of Rules

Perhaps the most difficult area to specify, the nature of rule enforcement is still considered important. The planning group suggests questions intended to get at the legitimacy the rules have with students, the frequency with which they are enforced, and the fairness of their application.

To what extent do you think the rules for student conduct are fair ones?

How often do students who misbehave in class get away with it?

How often are the rules in your school fairly and firmly enforced, regardless of who violates them?

Next Steps:

1. Resource Group to select indicators and measurement approaches based on priority importance, given by the Educational Goals Panel, the recommendations in this technical report, and any reporting constraints it sees.

    a. Identify existing items, refine items, or develop new items, and negotiate inclusion on appropriate instruments.

    b. Propose methods for improving state participation and quality of data across all measurement systems.

    c. Propose data shells for annual reporting.

2. Resource Group to finalize annual reporting plans.

APPENDIX A

Preliminary work of the Technical Planning Subgroups on Citizenship and The National Assessment of Educational Progress (NAEP).

GOAL 3 TECHNICAL PLANNING SUBGROUP ON CITIZENSHIP

TECHNICAL PLANNING SUBGROUP MEMBERSHIP

Leader		
	David Hornbeck	Education Advisor, Maryland
Members		
	Gordon Ambach	Council of Chief State School Officers, District of Columbia
	John Buchanan	Council for the Advancement of Citizenship, District of Columbia
	Todd Clark	Constitutional Rights Foundation, California
	Barbara Gomez	Council of Chief State School Officers, District of Columbia
	James Kielsmeier	Youth Leadership Council, Minnesota
	Lauren Resnick	University of Pittsburgh, Learning Research and Development Center, Pennsylvania
	Joan Schine	National Center for Service Learning in Early Adolescence, New York

REPORT OF THE GOAL 3 TECHNICAL PLANNING SUBGROUP ON CITIZENSHIP

Submitted to

The National Education Goals Panel

through

THE GOAL 3 RESOURCE GROUP ON CITIZENSHIP

for presentation at

The National Education Goals Panel Meeting

July 1, 1991

July 1, 1991

TO: THE NATIONAL EDUCATION GOALS PANEL

FROM: DAVID HORNBECK, LEADER
THE GOAL 3 TECHNICAL PLANNING SUBGROUP ON CITIZENSHIP

GOAL 3 TECHNICAL PLANNING SUBGROUP

STATEMENT ON THE TECHNICAL PLANNING SUBGROUP'S REPORT ON CITIZENSHIP

As you know, the President and the Governors identified as one purpose of goal 3 that all students "...may be prepared for responsible citizenship". In addition to the goal statement, one of the objectives of goal 3 is that, "All students will be involved in activities that promote and demonstrate good citizenship, community service, and personal responsibility."

Implicit in this commitment, I believe, is the notion that democracy is not sustained if it relies on its math/science prowess alone. It also requires a knowledgeable participation of the community in its own affairs. The sustenance of our freedom depends on active citizens. The schools have, from their inception, played an important role in the transmission of these democratic values.

After the last Goals Panel meeting, Dr. Forgione asked that I prepare a brief memorandum for the Panel regarding potential indicators related to citizenship. In the short time before the deadline, I consulted with Lauren Resnick and Gordon Ambach who also serve on Resource Group 3; John Buchanan, Council for the Advancement of Citizenship; Todd Clark, Constitutional Rights Foundation; Barbara Gomez, Council of Chief State School Officers; James Kielsmeier, Youth Leadership Council; Joan Schine, National Center for Service Learning in Early Adolescence; and members of the Goals Panel staff. I was also informed by previous conversations with other Resource Group 3 members. While the time frame has not allowed me to seek their concurrence with this memo, the recommendations in this memo are consistent with those conversations. The Panels' staff also concur with these recommendations.

Given the complexity and enormity of the task, both Resource Group 3 and the Panel have been properly focused on 1991 and long-term indicators of student achievement in the various discipline areas mentioned explicitly in the goals. However, it is also imperative, in my view,

that interim indicators for measuring citizenship be included in the 1991 report and that a process be put in place to identify and/or develop permanent indicators related to citizenship for the future.

Indicators for the September 1991 Progress Report:

There are presently three indicators which I recommend for 1991. According to the Panel staff, each has a solid data source available.

- Community Service -- In 1990, a nationally representative sample of 10th grade students were asked in the National Educational Longitudinal Survey (NELS) to respond to "Volunteering or performing community service" (rarely or never; less than once a week; once or twice a week; and every day or almost every day). The Panel staff believes this data may be available for the September 1991 report. We strongly urge that you seek it.

- Voter Registration of 18-20 Year Olds -- National data are available from the Census Bureaus' Current Population Survey regarding voter registration of 18-20 year olds in 1988. They are also available by race and gender. We urge that these data be included.

- Knowledge of Citizenship -- Our focus on service and voter registration underlines the view that citizenship is a call to action, not just passive knowledge. Nevertheless, civic education does have a content base involving the understanding of the institutions of government, how they operate, and how civic learning can transfer to a variety of problem contexts. NAEP has data available from its 1988 12th Grade Civics test. We recommend their inclusion in the 1991 Report.

While these three data sources are not what we ultimately hope to use, they are outcome measures and their inclusion will make clear that the President and Governors were serious in identifying citizenship as an important part of goal 3.

Future Progress Reports

The indicators recommended for 1991, as is the case with many of the indicators chosen for other goals, need to be strengthened. One obvious deficiency, also shared by other goal indicators, is the absence of state-by-state data. In addition, particularly in the community service and knowledge of citizenship areas, further review is important. With respect to community service, it will be important to consider what standards should apply. Should a one hour or one-time

effort "count"? Since our focus is on education, should a premium be placed on community service accompanied by reflection? There are also other quality control issues that should be addressed. With respect to civic education, should the national examination system that will be considered related to the other parts of goal 3, take civic education into account?

Resource Group 3 believes that citizenship is an important part of goal 3. While all of its members are more than productively occupied in a variety of ways, they have expressed their willingness to oversee the further examination of Citizenship indicators for permanent status because of the importance they attach to this goal. Thus, I urge you to ask them to do just that and report to the Panel no later than December, 1991.

GOAL 3 TECHNICAL PLANNING SUBGROUP ON THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP)

TECHNICAL PLANNING SUBGROUP MEMBERSHIP

Leader

Richard Mills	Vermont Department of Education

Members

R. Darrell Bock	University of Chicago, National Opinion Research Center, Illinois
David Cohen	Michigan State University
Paul Copes	Bloomfield School District, Connecticut
Robert Glaser	University of Pittsburgh, Learning Research and Development Center, Pennsylvania
David Haynes	St. Paul School District, Minnesota
Mark Musick	Southern Regional Education Board, Georgia
Thomas W. Paysant	San Diego City Schools, California
Suzanne Triplett	North Carolina State Department of Education
Leland Yee	San Francisco Board of Education, California

September 23, 1991

TO: THE NATIONAL EDUCATION GOALS PANEL

FROM: RICHARD MILLS, LEADER
TECHNICAL PLANNING SUBGROUP ON THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP)

GOAL 3 TECHNICAL PLANNING SUBGROUP

STATEMENT ON THE TECHNICAL PLANNING SUBGROUP'S CHARGE

This Technical Planning Subgroup is scheduled to meet in November 1991 to begin to develop recommendations for monitoring progress toward the National Educational Goals using the National Assessment of Educational Progress (NAEP). The group will submit a report to the Panel in Winter 1991 that will outline an interim and long-term data collection plan, including:

- subject matter areas and grade levels to be assessed;
- periodicity of the assessments;
- reporting at national and state levels;
- alignment of NAEP frameworks and test objectives to national content standards;
- use of performance assessment tasks;
- quality and breadth of items and tasks in each subject matter area;
- establishment of achievement levels; and
- measures of other relevant variables such as instructional practices.

NATIONAL EDUCATION GOALS PANEL

WRITE TO THE PANEL – ABOUT THE NATIONAL EDUCATION GOALS

PLEASE CIRCLE ALL THAT APPLY TO YOU. I AM A:

Parent Educator Business Person Concerned Citizen

___ I am writing in response to the expert recommendations the Panel received from the field.

___ I am writing in response to issues and/or questions raised by:

GOAL 1 ___, 2 ___, 3 ___, 4 ___, 5 ___, 6 ___.
(check appropriate space)

___ I am attaching comments of general concern about the National Education Goals.

___ I am attaching an idea or suggestion for future Goals Reports.

___ I am requesting more information about the National Education Goals.

Publication requested: _____

1991 Goals Report ___ 1991 Goals Report Executive Summary ___

Name:_____

Address:_____

City:_____

State: _____ ZIP:_____

Mail to: National Education Goals Panel
1850 M Street, N.W., Suite 270, Washington, DC 20036

NATIONAL EDUCATION GOALS PANEL

PUBLIC COMMENT FORM

Name:_____
(please print)

Address:_____

City:_____

State: _____ ZIP:_____

_____

COMMENTS:

Signature:_____Date: _____

_____

Mail to: National Education Goals Panel
1850 M Street, N.W., Suite 270
Washington, DC 20036

National Education Goals Panel

EXECUTIVE DIRECTOR

Pascal D. Forgione, Jr.

ASSOCIATE DIRECTOR FOR ANALYSIS AND REPORTING

Martin E. Orland

PROFESSIONAL STAFF

Edward J. Fuentes
Laura Lancaster
Leslie A. Lawrence
Cynthia D. Prince
David L. Stevenson
Charles J. Walter
Emily O. Wurtz

with assistance from

Carol Jay Stratoudakis

SUPPORT STAFF

Tia M. Cosey
Darius D. Dickens
Michael F. Smith
Edna E. Wilson